Hume. F.P.
1936

Measuring Teacher efficiency

BOSTON UNIVERSITY
SCHOOL OF EDUCATION
LIBRARY
Ed.
The Gift of Frederick P. Hume
Thesis
Hume
1936

ED
Stored
Hume
1936

BOSTON UNIVERSITY

SCHOOL OF EDUCATION

Thesis

MEASURING TEACHER EFFICIENCY
(WITH SPECIAL REFERENCE TO MASSACHUSETTS)

A SURVEY OF THE HISTORY, THEORIES, PURPOSE, AND EXTENT OF TEACHER RATING

Submitted by

FREDERICK PARKER HUME

(B.S. in ED., Boston, 1935)

In partial fulfillment of requirements for
the degree of Master of Education

1936

First reader : Roy O. Billett, Professor of Education
Second reader : Herbert Blair, Professor of Education
Third reader : Jesse B. Davis, Dean, School of Education

Boston University
School of Education
Library

TABLE OF CONTENTS

	Page
Chapter I: Value and Purpose of Measuring Teacher Efficiency	1
Aim of Study	1
Need for a Method	1
Scope of Study	3
Purpose of Measuring Teacher Merit	4
Need for Definition	6
Salary Schedule Based on Merit	7
Summary	11
Chapter II: A Brief History of Rating Teacher Merit	14
Definition of Terms	14
Initial Effort	14
Weighted Score Card	15
Boyce's Score Card	17
Self-improvement Score Card	18
Elementary Teachers' Test	19
The Indiana Method	20
McCall's Thesis	22
Summary	23
Chapter III: More Recent Concepts about Measuring Teacher Merit	26
Need for Standardization	26
Need for Group Judgment	28

Digitized by the Internet Archive
in 2016

https://archive.org/details/measuringteacher00hume

Chapter III:	Qualities of Merit Listed by Students	30
	Qualities Listed by Educators	33
	Knudsen's Survey	37
	Teacher Merit Based on Pupil-Achievement	40
	Ability Index Plan	41
	McCall's Achievement Quotient	43
	Merit Based on Teacher's Exercises	45
	Summary	48
Chapter IV:	Types of Rating Devices	51
	A Graphic Scale	51
	A Five-Point Method	53
	Descriptive Method	53
	Freyd's Scoring Stencil	54
	Central Tendency Plan	56
	Starrak's Device	59
	One-Hundred Point Plan	63
	Five-Hundred Point Plan	64
	Personal Inventory Method	65
	Drive and Response Method	66
	Correlated Study	67
	Summary	68
Chapter V:	Measuring Teacher Merit in Massachusetts	74
	Scope of Questionnaire	74
	Limitations of Study	74
	Source of Data	76

Chapter V: Criteria for Judging Candidates 80
Methods of Selecting Teachers 82
Rating Teachers in Service 83
Salary Schedules 86
Representative Methods for Rating Teachers 89
Summary 100

Chapter VI: Conclusions and Recommendations 104
Indications of the Study 104
Outstanding Weakness of Methods 105
The Providence Plan 108

Appendix 112

Bibliography 113

INDEX OF TABLES

		Page
Table 1	Proportions of Superintendents employing various devices in the selection and promotion of teachers	76
Table 2	Towns and Cities from which usuable replies were received	79
Table 3	Criteria used by 17 superintendents in judging prospective candidates	80
Table 4	Correlation between examination scores and teaching success	81
Table 5	Criteria employed by superintendents in judging merit of teachers in service	83
Table 6	Criteria on the basis of which salaries are increased as reported by 7 superintendents	87

Page

Plate I Map showing the cities and towns included in survey and from which replies were received 74-a

CHAPTER I

VALUE AND PURPOSE OF MEASURING TEACHER EFFICIENCY

Need for a Method

Aim of study.- Since the first part of the twentieth century efforts have been continually made by numerous educators to construct a valid, reliable, and objective method for measuring teacher efficiency. It is not the purpose here to establish such a method for others with more time and finances than the writer will be needed to solve the problem. The aim of this study is to so collate, summarize, and interpret the materials bearing on the subject as to offer a foundation for the further study and experimentation necessary to obtain a satisfactory result. The writer has avoided subjective evaluations of the various studies summarized. Obviously, the reader should as a rule attach more weight to those investigations which have the weight of much objective evidence behind them. The study also proposes to compare the present practises of Massachusetts school systems with criteria common to the general methods of measuring teacher efficiency, the information having been obtained by a questionnaire sent out to a limited number of superintendents.

Aim of chapter.- The chapter will attempt to show, through numerous quotations and excerpts, what concepts and attitudes educators hold concerning the value, need, and purpose of a method for measuring teacher efficiency, and to show how controversial is the problem.

Need for a method.- There is a need for a method of measuring teacher efficiency when we consider the problems related to the improvement of teachers in service, to the determination of teacher promotion, or dismissal, or selection, or to the adjusting of the salary schedule

based on teacher merit. Educational literature abounds with schemes and methods for administrative purposes but they are too subjective and unreliable. In referring to the need of a scientific method, Bossing says, "It is quite evident....that a careful analysis of our criterion has not been a matter of concern by students of this subject." 1/

Perhaps there can be no really objective method of measuring teacher efficiency when human judgment is called into play, but we should have a means of evaluating the qualities which constitute teacher merit that are as near to an impersonal character as it is possible to obtain.

Engleman says by way of encouragement:

> "I think there is no reason for discouragement because most rating schemes today are largely subjective. Every capable and skilled critic or supervisor succeeds in utilizing some sort of a scale or rating system as a standard of excellence. Standards that are admittedly subjective, work. They have long done so,....While we must supplement them as rapidly as we know how to do it with something perhaps more reliable because more objective, we need to remind ourselves that a large part of the best of any teacher's influence and ministry cannot be measured successfully by any sort of standard, subjective or objective, until years after the teaching is done." 2/

While it may be conceded that the best part of any teacher's worth is measured by the growth of pupil achievement in subject matter, yet, other qualities of a more subtle nature which involve

1/ Nelson L. Bossing, "Teacher-Aptitude Tests and Teacher Selection", Research in Higher Education, Bulletin No. 12, 1931, p. 121.

2/ J. O. Engleman, "The Place of Objective Standards in Judging Student Teaching", Educational Administration and Supervision, 17:328, May 1931.

social as well as academic growth need to be considered from both the teacher and pupil viewpoint. The classroom situation must be the focal point of any study dealing with teacher merit. It is here that many of the variables which influence a judgment may be controlled while that which occurs outside the classroom is relatively invalid and uncontrollable.

Scope of study.--This study is primarily interested in judging validly, reliably, and objectively the qualities of merit of teachers in service, yet, it must consider also those qualities by which teachers are selected for their initial work. Literature of recent date upon this subject of measuring efficiency is not too abundant and new ideas of a constructive, scientific nature are seldom proposed. At present the trend of educators is to treat measuring methods only for supervisory purposes. The validity of such findings, in the writer's mind, even for supervision is questionable for no valid, reliable, or objective method has been produced. The following comment made in the Research Bulletin of the National Education Association shows that some importance is attached to measuring teacher efficiency in service.

> "Additional study of present practices in locating, selecting, and appointing teachers may not be needed, but it seems evident that studies to ascertain the best criteria for precluding....and for appraising the work of teachers already in service should be undertaken by some educational agency. Several questions need to be answered. Among these are: (1) What constitutes teaching success? (2) How may the success of teachers in service be measured? When these questions are answered, the problem of teacher selection will be well on its way to solution." 1/

1/ "Administrative Practices Affecting Classroom Teachers", National Survey of Secondary Education, 10:115, March, 1932.

Although this comment indicates the need for further study in order to obtain criteria for measuring the merit of teachers already employed, yet no mention is made of the reason or purpose of such a method.

Purpose of measuring teacher merit.--Educators do not agree that the purpose of measuring teacher merit is for promotion, retention, or salary-schedule adjustment. The current opinion of textbooks and educators is that measurements should be used as a supervisory function for diagnosing and correcting the faults and difficulties of teachers and teaching technique. Weber says the only legitimate purpose of any rating scale is to further the maintenance of an attitude of co-operative endeavor and he mentions, "Still 43 per cent of the cities have some kind of formal rating scale but only 10 per cent use such a scheme as a basis for a differentiation in salary."1/

This statement was recorded two years after the Research Bulletin of the National Education Association stated, "A smaller or at least no larger per cent of cities now rate teachers than did three years ago."2/ The contradiction implied by these two statements would indicate that interest in measuring teacher merit is still maintained. An editorial of a more recent date, in speaking about those teachers who ought to be removed because they are a hindrance to a school system, says:

1/ Oscar F. Weber, "Problems in Public School Administration", New York, The Century Company, 1930, p. 259.

2/ Ibid.

> "The solution of the problem it would seem, lies in the modernization and legal acceptance of periodic appraisals of the efficiency and professional spirit of teachers. Such appraisals must include not merely the judgment of principals and superintendents on the personality and general interest of teachers. It may well include the directly measurable teaching results observed in passing marks, results of achievement tests,....."1/

It is fair to assume that if a valid and reliable method could be devised, more emphasis would be placed upon using the results of such a method as a basis for salary adjustment, promotion, and dismissal of teachers. The problem of measuring teacher efficiency is still an active one although not many educators are doing active work in this field. The recent Kappa Delta Pi research publication2/ and the following quotation indicates fairly conclusively that the problem still is unsolved and that the trend is towards the use of measuring efficiency for the supervision and improvement of teachers.

> "The results of supervisory tests given periodically for the purpose of checking the efficiency of the teacher's instruction should be revealed to the classroom teacher in terms of specific suggestions for the further improvement of the situation."3/

While the present purpose of measuring teacher efficiency has been pointed out, yet this purpose will be frustrated if confusion as to the meaning of the terms or items used in a method is present.

1/ "Shall the Teacher-Tenure Laws be Amended?", American School Board Journal, 90:43, May 1935.

2/ W. H. Lancelot, A. S. Barr, T. L. Torgerson, et al, The Measurement of Teaching Efficiency, The MacMillan Co, New York, 1935, p.139.

3/ Harry A. Green and Albert N. Jorgenson, The Use and Interpretation of High School Tests, Longmans, Green and Company, N. Y., 1936, p.582.

Need for definition.--Sorenson, in his study of intelligence and teaching success, found that the lack of correlation between the two items was caused by the unreliability due to lack of definite standards and the lack of defining subtle, intangible, and elusive traits such as co-operation and enthusiasm.[1] One reason for the unreliability of rating scales has been this personal factor of subjective judgment. Until variables such as this one in which almost any interpretation can be given as to what enthusiasm really is or consists of, depending upon the judge, no valid or reliable results will be obtained. When it is possible for enthusiasm to imply happiness, zealousness, confidence, energy, leadership, or any other meaning which the individual cares to read into the interpretation, it is not surprising that not much progress has been shown in this field.

Furfey claims that the analyzed components of a general trait become more reliable when the trait is broken up into a number of specific factors,[2] and Williams says:

> "If ratings are to be reliable, definitions of items should be in terms of what a person does. The first step in the diagnosis is that of discovery; therefore, the initial emphasis should be on description rather than on interpretation or judgment."[3]

1/ Herbert Sorenson, "Why Teaching Success Does Not Correlate Highly with Measured Intelligence", Educational Administration and Supervision, 15:602-06, Nov. 1929.

2/ P. H. Furfey, "An Improved Rating Scale Technique", Journal of Educational Psychology, 17:45-8, Jan. 1926.

3/ Eula S. Williams, "A Personality Rating Form for Elementary-School Pupils", Elementary School Journal, 34:23, Sept. 1933.

It is evident that definition of qualities so as to control as many subjective factors as possible should be one of the fundamental steps in the construction of a valid, objective, and reliable method. Some scientific method would be necessary if public education based on a civil service idea as advocated by Lewis should ever take place.[1] Such a method could also be made the basis for salary increments.

Use of a Valid and Reliable Method

Salary Schedule Based on Merit.--The only scientific approach to the creation of a salary-schedule seems to be by means of measuring teacher efficiency. If a reliable method for such a purpose could be established, much of the difficulty in making the adjustments which are necessary by superintendents in order to retain or dismiss their teachers would be avoided. A salary-schedule based on valid, reliable, and objective data concerning a teacher's ability or worth to a school and community would eliminate to a great degree the political influences so prevalent in the school systems today. Many systems boast of a schedule based on teacher merit but the criterion of this merit is usually experience and training. Automatic salary increases are the rule rather than the exception and a teacher usually is advanced a fixed sum for every year of service up to a specified maximum amount. How efficient a teacher may be is uncertain but studies claim that the

1/ E. E. Lewis, "Desirable Principles of Teachers' Salary Schedules" American School Board Journal, 80:45, May 1930.

maximum period for reaching a high degree of teaching efficiency is from six to eight years. With the automatic salary increase based on experience and length of service, no adequate or scientific knowledge can be ascertained concerning a teacher's efficiency.

Experience and training are two criteria for judging a teacher's worth but how much more scientific could a salary schedule be which not only considered these two items but also included as its basis a reliable measure of teaching efficiency. It would seem that teaching ability would not stagnate at the end of a decade of work under such conditions for those not professionally inclined would be dismissed from their positions. With such data to present to the school committee and community it might be possible to eventually prove that the teaching occupation is entitled to salaries comparable to the other professions. It is worth the effort, and a valid, reliable, and objective method for measuring teacher efficiency seems to be the point of attack, towards which these educators should turn who desire a schedule based on scientific principles.

The remarks of several educators indicate that merit is to be considered in salary schedule adjustments.

Newlon say, "We must in the near future work out a practicable plan of paying teachers on the basis of merit. Up to the present time, we have not solved that problem."1/

1/ Jesse H. Newlon, "Salary Differences, If Any, Based on the Rank or Class of the Teacher", Addresses and Proceedings of the National Education Association, 59:841, 1921.

Corson has the same attitude when he states, "Demonstrated personal power combined with scholarship constitutes the right basis for salary reward." 1/

Goslin says:

> "It seems that the pay which teachers receive should be dependent upon their degree of success....The only answer is ....that no fair, accurate, and satisfying method has been determined for measuring a teacher's success. Until some accurate means of measuring has been devised,it is futile to attempt to unite a teacher-rating plan with a salary-schedule." 2/

Ruediger, judged by his comment, is against basing a salary increase upon teacher-rating:

> "The administration of a salary-schedule, if it is to induce and maintain a wholesome morale, must operate impersonally and to a great extent automatically. It must rest on principle rather than on an annual recurrence of personal judgment, no matter how fair this judgment aims to be." 3/

A spirit of co-operation between teachers and superintendents is implied by Lindsay when he favors a salary schedule based on a measurement of merit. He says:

> "But the estimation of the teacher's value to the district is the ultimate key to the salary situation.It cannot be neglected. A sound salary policy cannot be built without it. The school administrator ....if he does not care, or thinks it inadvisable at any given time to face the situation, he may as well leave the salary question alone and allow salary determinations to proceed along the old haphazard, unbusinesslike, unscientific, wasteful and inefficient channels.

1/ David B. Corson, "The Relation of Scholarship and Growth in Scholarship to Merit and Pay", Addresses and Proceedings of the National Education Association, 59:838, 1921.

2/ H. Goslin, "A Salary Schedule for Teachers in a City of Forty Thousand", American School Board Journal, 69:37, Oct. 1924.

3/ W. C. Ruediger, "Rating Teachers", School and Society, 20:266 August 30, 1924.

> channels. If a teacher-rating scale is to be adopted, the teachers of the system should know on what points they are to be judged, and what weight the points are to have. The points and weighting are wrong unless they present a detailed measure of a good teacher. Co-operation is a real reason for superintendent or board making known the methods by which their teachers are judged. Efficient salary schedule adjustments demand the use of rating scales of success in teaching. " 1/

It is interesting to note that Lindsay believed that the teachers should know on what basis they were judged and also what value was assigned to that judgment. A suggestion by the First Yearbook of the National Education Association that, "After all, it is the position and not the person which should determine the salary." 2/ is potent with possibilities but is limited, in the writer's opinion, because no definite criteria for judging the economic value of a position is available. The position of headmaster, for example, may pay a large salary in one school system and a small salary in another. The positions are identical but the economic factor (variable) influences the salary. No progress can be made along this line until a scientific basis for rewarding teacher merit is created so that the association of merit with salary is valid and reliable.

1/ E. E. Lindsay, Problems in School Administration, New York, The MacMillan Company, 1928, pp.187-251.

2/ "How the Single Salary Schedule Appeals to the Teacher," Department of Classroom Teachers, p.49, July 1926.

Summary

The different concepts of various educators in regard to the measuring of teacher efficiency have been presented. The fact has also been mentioned that a comparison will be made in a later chapter between the criteria common to the general methods for measuring teacher merit and data furnished by some Massachusetts superintendents in their replies to a questionnaire sent out by the writer.

Bossing has claimed that a careful analysis of the criteria for judging teacher efficiency has been neglected, and Engleman has expressed the opinion that the best results of a teacher's labor cannot be measured successfully until years after the teaching act. The writer, while conceding that the best part of a teacher's worth should be measured in terms of pupil achievement, and while he does recognize that other factors of a more social nature ought to be considered, yet he believes the focal point of any study for the purpose of devising a method for measuring teacher efficiency should be in the classroom situation. What occurs outside is relatively invalid.

The scope of the study has been limited principally to the measurement of teacher efficiency in service, while the aim is to summarize and interpret such material on the subject as the writer believes will perhaps be of some value for further study and experimentation.

In the course of the chapter, the 1928 Research Bulletin of the National Education Association has been quoted as raising the

baffling question, "What constitutes teaching success?", and it has also advocated that studies should be undertaken to ascertain the best criteria for judging teachers in service. Weber claims the only legitimate purpose of a rating scale is to increase the spirit of co-operation, and that only 10 per cent of those cities having some type of rating method use it as a basis for salary differences. Reference has been made to two recent publications, one by the Kappa Delta Pi society and the other by the co-authors Green and Jorgenson, showing that the problem of how to establish a valid and reliable method for measuring a teacher's efficiency still remains unsolved, and that the present trend is toward the use of measuring devices for supervisory purposes.

Sorenson found that the reason for no correlation between intelligence and teaching success has been the inability of the one doing the judging to grasp the meaning or concept of what constitutes such subtle traits as co-operation, enthusiasm and others of an intangible nature. The inferred conclusion is that all the items or qualities on a measuring device must be so defined that the variable of personal interpretation will not be read into the meaning. Furfey found that the analyzed components of a general trait are more reliable when the trait is broken into specific factors, and Newlon, Corson, Goslin, Lindsay, and Ruediger have been quoted to show that a salary schedule based upon the results of measuring teacher merit is plausible even if not possible under existing circumstances.

The First Yearbook of the National Education Association has suggested that the position command the salary instead of the man, but the writer believes the primary need is for a valid and reliable method for determining teacher merit before any relation between position and salary can be scientifically established. Although it is impossible to definitely settle here, or to come to any conclusions as to which course or group of ideas should be followed, nevertheless it is apparent that the need for a valid, reliable, and objective method still exists and that a salary schedule based on merit cannot be scientific until such a method is devised. The following chapter will give a brief review of what others have done in the earlier years of the movement.

CHAPTER II

A BRIEF HISTORY OF RATING TEACHER MERIT

Methods Before 1915

Definition of terms.- The three terms of ability, merit and efficiency will be used at various times throughout this study with the meaning being identical. They are indicative of the general nature or attributes of a teacher judged to be successful in the profession. The degree of teacher success, of course, is variable and relative.

The earlier administrators who were interested in knowing what their teachers were doing or were capable of doing, made no practical use of their findings for administrative purposes. It was not until the movement had become popular that educators with a professional spirit for their work began to consider teacher-rating as a basis for supervision, promotion, or dismissal of their teaching staff. Even then opinion was divided as to the value of such practises,and there still exists today the belief that such methods are too uncertain and unreliable for other than supervisory, self-improvement purposes.

The premier contribution in the field of teacher rating was by Meriam.

Initial effort.- Meriam's work in this field is considered to be the first major attempt at solving the problem of measuring a teacher's merit. The estimates of principals were taken as the best available mark for measuring efficiency and the marks of scholarship and practise teaching were taken from the school records of all the graduates of five normal schools from 1898-1902.

He did not attempt to make a score card nor did he use his findings in any administrative capacity. His criteria was based on : 1/

1. Scholarship during the normal school course
2. Rank in practise teaching during that course
3. Length of teaching experience

After Meriam's contribution, interest in rating teachers apparently lay dormant for five years for nothing more was accomplished until Eliot's plan appeared in 1910.

Weighted score card.- Eliot compiled a score card with minimal values assigned to the several qualities which were grouped under the various headings as: 2/

1. Physical Efficiency	Total-	12 points
a. Impressions-general		2 "
b. Health,etc.		2 "
2. Moral-Native Efficiency	"	14 "
a. Self-control		2 "
etc.		
3. Administrative Efficiency	"	10 "
a. Initiative, etc.		2 "
4. Dynamic Efficiency	"	24 "
a. Preparation, etc.		4 "
5. Projected Efficiency		6 "
a. Continuing preparation,etc.		2 "
6. Achieved Efficiency		24 "
a. Achievement, etc.		8 "
7. Social Efficiency		10 "
a. Intramural interests,etc.		2 "

1/ E. E. Lewis, Personnel Problems of the Teaching Staff", New York, The Century Company, 1925, p. 192.

2/ "The Supervision of City Schools", Twelfth Yearbook of the National Society for the Study of Education, Bloomington, Illinois, The Public School Publishing Company, 1923, Part 1, pp.68-69.

This schedule indicates a practical belief that there is a possibility of drawing up plans and criteria which would show those elements and qualities that are considered necessary in the equipment of a teacher. It is the forerunner of some present-day rating-scales which attempt to define each item of the scale so that a minimum of subjectivity or personal bias will be present to influence the judgment.

About this time two other educators made a contribution. Ruediger and Strayer, in attempting to answer the questions, "What are the fundamental qualities of teachers that should receive scientific and administrative attention?" and, "What is their relative importance and how may they be determined?", based their decisions on the following list of criteria:1/

Education
Years of Teaching Experience
General teaching merit
Health
Initiative
Personality
Teaching skill
Ability to carry out suggestions
Accord between teacher and pupil
Progressive scholarship
Social factors outside the classroom

A little later Hughes enlarged upon the schemes of his predecessors by adding more items to the list upon which teachers were to be rated;2/ but it was not until 1915 that another contribution, an outstanding one of its kind, was submitted.

1/ E. E. Lewis, Op. Cit. p. 193.

2/ Ibid, p.194.

A definite purpose.--Boyce formulated a score card for rating teachers which covered 45 different qualities grouped under five main headings as:[1]

Personal Equipment
Social "
Professional Equipment
Technique of Teaching
Results

He hoped to eliminate the rating of teachers on snap judgment, personal preference, comparison, political influence, and other factors of a subjective nature and he believed his method was a fairly scientific one for determining the promotion and vocational guidance of teachers. He also used a five point plan as: Very poor, poor, medium, good, or excellent, for forming a judgment in respect to each of the forty-five qualities listed.

This plan, created with a definite purpose in view, shows that the problem of measuring teacher efficiency had become a worthwhile one ten years after its inception by Meriam. At that time, and even today, much emphasis is placed upon measuring the personal attributes of a teacher rather than the results of her teaching. A different plan, however, was presented in 1916.

Methods After 1915

"Results of teaching" plan.--Connor[2] devised a method based on the following principles:[2] (1) Teaching, not teachers, should

1/ A. C. Boyce, "Methods of Measuring Teachers' Efficiency"; Fourteenth Yearbook of the National Society for the Study of Education, Chicago, Univ. of Chicago Press, 1915, Part II.

2/ W. L. Connor, "A New Method of Rating Teachers", Journal of Educational Research, 1:338, May 1920.

be rated, (2) Teaching is rated in terms of results only, (3) Results may be subjectively or objectively graded as inferior, below average, average, above average, or superior, but only in reference to the amount of improvement made and the ability to improve acquired by the particular group. He grouped the results of teaching under the following heads which are defined and given in detail in the original plan. The main headings are:

1. Thinking
2. Knowledge and skill
3. Initiative
4. Morale
5. Emotional reaction
6. Effectual self-control in situations socially significant
7. Deportment

It may be assumed, without any valid evidence, that the more recent concepts which have considered measuring teacher's ability upon pupil achievement may have owed their originality to this plan. Another method was produced by Rugg four years later with still a different purpose in mind.

Self-improvement score card.- Rugg made a score card in 1920 which consisted of two forms , A and B.[1] Form A has questions which appertain to the degree to which a teacher possesses certain qualities or has carried out certain activities. It is intended fundamentally as a means of self-improvement through self-rating.

His form B.is a device for assigning numerical ratings to teachers based upon a teacher-to-teacher or man-to-man comparison.

1/ H. O. Rugg, "Self-Improvement of Teachers through Self-Rating; A New Scale for Rating Teachers' Efficiency", Elementary School Journal, 19:670, May 1920.

The very best teacher whom one knows is used as the highest standard for purposes of comparison; the lowest standard is that of the poorest teacher. Rugg groups his questions under five heads as:

1. Skill in teaching
2. Skill in the mechanics of managing a class
3. Team-work qualities
4. Qualities of growth and keeping up-to-date
5. Personal and Social qualities.

He gives the best teacher rated a numerical value of 38, the better than average teacher a 32, and the poorest teacher a value of 6. An example is given of only one item.

1. Skill in teaching

Best teacher	38
Better than average	32
Average	30
Poorer than average	14
Poorest teacher	6

This method is similar to Eliot's in that numerical values are given to the ratings which enable the rater to score his results easily. The validity of such a method is questionable when based upon subjective interpretations as the man-to-man principle indicates.

A device similar to many of our present day ones for testing the psychological and professional knowledge of teachers was brought out two years later by Knight and Franzen.

Elementary teachers' test.--Knight and Franzen in 1922 devised a test for elementary school teachers which consisted of seven pages of questions grouped under 36 divisions.[1] These questions were

1/ E. E. Lewis, Op. Cit. p.198.

designed to test the professional knowlege which the teacher possessed. One example is given as follows:

"If a question is asked by the pupil that you don't know how to answer, what would you do? Put a cross before the one of these statements which best describes what you would do".

1. Tell the child you don't know.
2. Give him any answer that comes into your head.
3. Tell him to wait a minute and then look it up.
4. Go and ask the principal.
5. Tell him to look it up.
6. Tell him it's a foolish question.

Such a test may be desirable for self-diagnosis but it would be too cumbersome for actual classroom use. It is not a score card and it is given here to show how educators have diversified their methods in the attempt to help out the teaching situation. The last plan to be presented is in the form of a score card based on a percentage basis.

The Indiana Method.--The Indiana certification of teachers with experience is based on a score card dealing with qualities conferred by training and a measure of the native qualities of personality possessed by the teacher.[1] It consists of the following items based on a total value of 100 per cent.

A. Personality	20
1. Physical	
2. Mental	
B. As a student	15
1. Lines of study pursued	
2. Lectures attended	
3. Vacation schools attended	

1/ "The Supervision of City Schools", Op.Cit.p.66.

C. Professional Development	15
D. As an instructor 1. Preparation 2. Presentation 3. Comparison of interpretation based on childrens' experiences 4. Generalizations 5. Application	20
E. Government	15
F. Community interest	15

It might be said in passing, that item "D. bears a strong resemblance to Morrison's theory of teaching technique. It can be seen that all the methods which have been briefly reviewed are generally of an individualistic nature. They represent different ideas of what educators believe a score card or other device for measuring teacher efficiency should be. There are no norms or mean based upon a nation-wide study, nor does there seem to be a reliable method devised for scientifically weighing items of quality. No one has as yet determined just how much of an influence any particular trait or characteristic has upon the success of a teacher. Dr. Chassel[1] attempted such an effort in her work at Columbia University but found no high correlations from which she could draw a valid conclusion. The trend at present seems to be for measuring teacher efficiency through the results produced in pupil behavior and achievement.

1/ G. T. Somers, "Pedagogical Prognosis: Predicting the Success of Prospective Teachers", New York, Teachers College, Columbia University, Contribution to Education, No. 140, 1923, pp. 68-69.

Lewis says:

> "In my judgment, the most important problem before the teaching profession today is the definition of what good teaching is in terms of the effects to be produced in children and the development of standard tests of teaching ability based on these definitions."1/

The teaching act offers a more objective basis, together with pupil achievement, than does any attempt to estimate teacher evaluation based on personal qualities. McCall has formulated a few principles for rating teachers with such a concept in mind.

McCall's thesis.--McCall has summarized what he believes are fundamental assumptions underlying a scientific procedure for rating and promoting teachers. Desirable changes in the pupils is the basis upon which he would judge his teachers' efficiency. What these DESIRABLE changes are or should be is not mentioned. Apparently administrators are expected to form their own opinions as to what constitutes "desirable changes", and this is probably one reason why such an atmosphere of vagueness surrounds the whole field of measuring teacher merit. McCall's theory is as follows:

1. The pupil is the center of gravity....of the educational system.
2. All the desirable paraphernalia of education exists for just one purpose, to make desirable changes in pupils.
3. The worth of these...,can be measured in just one way, by determining how many desirable changes they can make in pupils.
4. Hence the only just basis for selecting and promoting teachers is the changes made in pupils.
5. Teachers are at present selected and promoted primarily on the basis of their attributes, such as intelligence, personality, physical appearance, voice, ability in penmanship, and the like.

1/ E. E. Lewis, Op. Cit. p.204.

6. No one has demonstrated just what causal relationship, if any, exists between possession of these various attributes and desirable changes in pupils.
7. Scientific measurement itself is fair only when we measure the amount of desirable change produced in pupils by a given teacher. The measurement....requires both initial and final tests.
8. Scientific measurement is fair only when we measure amount of change produced in a standard time.
9. Scientific measurement is fair only when we measure amount of change produced in standard pupils.
10. Scientific measurement is fair only when the measurement is complete. Absolute completeness would require a measurement of the amount of changes made in children' purposes as well as their abilities.1/

SUMMARY

The theories and plans for measuring teacher efficiency listed in this chapter are only a few of the more prominent attempts that have sought to solve the problem. While these plans have not proven their validity, reliability, or objectivity yet they have done much toward crystallizing thought on the subject, and to prove to educators that the answer as to how the efficiency of a teacher can be measured is still unknown. The methods which have been briefly reviewed are illustrative.

Since the early part of the twentieth century, educators have been contributing plans for measuring teacher efficiency. The pioneer attempt was by Meriam in 1905. He used no score card and based his judgment on scholarship, practice teaching, and experience.

Eliot, however, compiled a card with values (points) assigned to the various qualities. Ruediger and Strayer, Hughes, and Boyce

1/ William A. McCall, "How to Measure in Education", New York, The MacMillan Company, 1927, p.152.

were later contributors, with Boyce formulating a score card containing forty-five items considered as necessary qualifications for teaching. He used a five-point scale described as "from very poor" to "excellent" and the purpose of his plan was to serve as a basis for the promotion and vocational guidance of teachers. This plan was outstanding in 1915.

In 1916, Connor followed Boyce, using also a five-point scale and emphasizing the fact that teaching and not teachers should be rated. Rugg produced his double form score card in 1920 based on the purpose of self-improvement; he weighed the items similar to Eliot and used the "man to man" basis for making his measurements.

Knight and Franzen contributed a test for elementary school teachers in 1922 designed to ascertain the professional knowledge of teachers, and a year later the Indiana method based on training and personality qualities with the items weighted on a percentage basis was produced.

McCall's theory based upon pupil achievement measured through scientific procedures is also mentioned. He stresses the need for controlling such experimental variables as: teachers, time, pupils, for giving initial and final tests, and for complete measurement.

Educators have recognized the fact that there should be some systematic method for evaluating teacher efficiency. Whether the purpose of such measurement is to improve teachers in service; to serve as a basis for promotion or dismissal; or to give merit as the criteria for salary schedule adjustment....this does not matter.... the need is obvious.

Interest in measuring teacher efficiency did not increase for five years after Meriam's initial contribution and Eliot's method was probably the forerunner of our present day rating scales which attempt to define their items so as to eliminate ambiguity and subjective interpretation. The personal attributes and the results of pupil achievement have been considered as valid criteria for judging teacher ability and the writer has questioned the validity of a method such as Rugg's when based upon the subjective, man-to-man principle. Reference has also been made to Dr. Chassel's work to show that no-one has determined the influence of any particular trait upon the teacher's success.

Most educators should realize that there must be a definite knowledge of the constituents of good teaching before any criteria that may be considered valid as a means of measuring teacher merit can be set up. With this in mind, therefore, the following chapter will discuss those elements, items, or qualities of merit that have been considered, both in theory and practice, by people who have been actively engaged in the task, as absolutely necessary for teaching success. No evidence is shown that these earlier workers in this field ever based a salary schedule on the results of their findings.

If a reliable, valid and objective method could be constructed, school administrators would have no further difficulties in the adjusting of teachers' salaries.

CHAPTER III

MORE RECENT CONCEPTS ABOUT MEASURING TEACHER MERIT

Standardization Necessary

Purpose of chapter.- The previous chapters have shown the purpose and need for a method of measuring teacher efficiency. Some of the more recent ideas and theories will now be presented so that a definite knowledge of the basic criteria necessary for the formulation of a valid, reliable, and objective method may be obtained. The writer realizes the magnitude of the problem and his own limitations in attempting to present all the factors involved in such a study, but it is hoped that at least the major factors will illustrate sufficiently the fundamental principles.

Need for standardization.- Opinions as to what items are essential for teaching success have been gathered together in many studies that have sought a basic criteria for judging efficient teaching. These opinions have coincided in respect to some quality of good teaching quite frequently, but the importance of such qualities has been assigned different grades, values, or weight. Betts gives an explanation as:

> "Sooner or later, in most of the controversial issues regarding the education of a teacher, the question is asked, 'How effective is this teacher?' As often as the question is asked, it is answered, but no two people give the same answer. The philosophy of the person doing the rating, acquaintance with the teacher and other factors influence the estimate. Each proponent of a different plan or program of teacher education has a different answer. Such confusion in thinking causes much educational effort to be at cross-purposes. Clarity of thinking is much needed so that concerted effort may be directed toward a solution of this important problem. By what scales are persons measuring the effectiveness of teaching? The answer

> answer at present seems to be that each measurer has his own unique measuring stick. The repute in which a teacher is held by one measurer may be entirely different from that of another, and a reputation for being an effective teacher may be quite independent of technical skill as a teacher."[1]

Individual effort in this field should not be condoned because it has led to confusion in thinking but it ought to function on principles scientifically established from controlled experimentation. The two following quotations serve to emphasize the need of a standard.

Almy says, "One of the most potent reasons for unreliability of teacher ratings is the failure to standardize the steps of the rating devices used."[2]

The 1928 Bulletin of the School of Education of Indiana University contributes:

> "Experiment has revealed the fact that if the measures of these less tangible behavior traits even approach reliability and validity, at least these conditions must be met: (1) the rater, the one who estimates the amount of a given trait in another, of these behavior characteristics, must define and use some sort of a fixed standard by means of which to compare the behavior observed....."[3]

Hughes sets forth some general principles illustrating quite well the need for standardization and definition of items if a measuring instrument is to be obtained for either subjective characteristics of a teacher's personality or objective facts in teaching technique. His principles are:

1/ G. L. Betts, "Philosophy in the Measurement of Teaching Ability", School Life, 17:168, May 1932

2/ H. C. Almy, "A Teacher-Rating Scale of Determined Reliability and Validity", Educational Administration and Supervision, 10:182, Mar. 1930

3/ "A Method for Studying the Unmeasured Traits of an Individual", 4:66, July 1928

1. Unity of definition....Every item in the rating scale must be defined clearly.
2. Behavioristic definiton....Every item should be defined in terms of behavior.
3. Relation to situation....Opportunity must be present to exercise the desired trait.
4. Relativity in rating....Comparisons from a standard.1/

As the automobile industry has standardization throughout its system, so should our educational systems have such standardization in their methods for measuring teacher merit. Definition of the items or qualities of merit are an integral part of the standardization process for individual opinion undefined is unsound.

Need for Group Judgment.--When Leonardo DaVinci was painting his Battle of Anghiari on the wall of the Council Hall in Florence the head of the Republic complained that the artist had made war look revolting. Leonardo proposed that the citizens decide by vote whether his picture was good or bad.

> "Firstly, because the certainty is mathematical--you have only to count the votes to know the truth. Secondly, every intelligent man, if he be alone, is apt to fall into error; whereas ten or twenty thousand fallible mortals convened together cannot make any mistake."2/

Here, in the fifteenth century, is evidence that combined group judgment is the only valid method of estimating the merit of a thing. Opinions of more recent date are also consistent in the belief that more than one rater should measure teacher merit. Just how many people should be involved is not known nor definitely stated.

1/ W. Harding Hughes, "General Principles and Results of Rating Trait Characteristics", Journal of Educational Method, 4:421-32, June 1925.

2/ Editorial--"Against Stupidity", Boston Sunday Globe, March 8, 1936 page 4.

because educators have no scientific basis for defending their opinions. The principle of group judgment applies, however, both to the individual cases where a teacher is being rated and also to a study needed today to determine definitely what criteria should constitute a standard method for measuring teacher efficiency.

Jacobs mentions that: "The most reliable criterion at the present time....is the conscientious and deliberate opinion of competent judges."1/

Strebel says:

> "There are no adequate objective measures of philosophic teaching. Judgment of results must be wholly subjective and therefore subject to wide fluctuations of opinion. The commonly accepted safeguard to be used under such conditions is the check afforded by group judgment."2/

A study by Towner of recent date, found that observation and judgments of superintendents and their assistants were the two criteria chiefly used in fifty-three cities to determine the evaluation of principals.3/

It thus appears that there must be a definite standardized method for measuring teacher efficiency and that such a method must be based not upon any individualistic measure but upon group judgment.

1/ C. L. Jacobs, "The Relation of the Teacher's Education to Her Effectiveness", New York, Teachers College, Columbia University, Contribution to Education, No. 277, 1928, p.25.

2/ R. F. Strebel, "Purposes of Evaluation of Student Teaching"; Educational Administration and Supervision, 17:340, May 1931.

3/ Earl M. Towner, "The National Elementary Principal", Bulletin of the Department of Elementary School Principals of National Education Association, 15:14-16, Feb. 1936.

The criteria upon which group judgment is to be exercised for measuring teacher ability is unlimited and confusing when they are established on general concepts held by individual educators. Barr and Emans found in their study of 209 rating scales that a great variety of items was used to characterize teaching and teaching ability; that these items were undefined, and that the social and personal traits of a teacher were mentioned more frequently than all the other traits.[1/]

Such a diversity of opinion, lack of knowing just what a term means, and the biased selection of social and personal traits as implied in the above report are reasons why a different attack must be made on this question of measuring efficiency if any further progress is to be made in this field. In the writer's opinion, delimitation of rating scale items by means of definition and common understanding brought about through a national experiment under central control should be the initial step. What some of these items may be which will constitute the criteria deemed essential will be illustrated by samplings of no definite validity or reliability, and limited both as to the number of items and to the number of examples.

Constituent Qualities

Qualities of Merit Listed by Students.--From time to time studies have been made with the assistance of the student body of an institution as to the qualities of a good teacher. A few examples

1/ A. S. Barr and Lester Emans, "What Qualities are Prerequisites to Success in Teaching?", Nations Schools, 6:60-64, Sept. 1930.

of such efforts will not be out of place here. Although a variety of opinions will be expressed, yet it is noticeable that some of the items are similar.

Newmark obtained the following list from 223 pupils.[1]

Best teachers	Poorest
1. Skillful in 'Getting ideas across	Weak in 'getting etc.'
2. Human	Lacks self-control
3. Impartail	Partial
4. Good disciplinarian	Poor disciplinarian
5. Sympathetic	Lacks common sense

Flory based his findings on the reports of 370 college students. Some of his items are similar to those of Newmark but discrepancy of opinion is apparent.[2]

Best teachers	Poorest
1. Impariality	Carelessness
2. Sympathy	Impatience
3. Appearance	Stubborness
4. Patience	Discourteous
5. Interest in extra-curricula events	Superiority complex

Macdonald obtained the following from 320 students. Not one of her items is similar to the above examples.[3]

1. Sense of humor	Injustice
2. Justice	Lack of consideration
3. Cosideration	Poor voice and address
4. Friendliness	Prejudice
5. Appropriate dress	Poor appearance

1/ David Newmark, "Students' Opinions of their Best and Poorest Teachers", Elementary School Journal, 29:576, April 1929.

2/ Charles D. Flory, "Personality Ratings of Prospective Teachers", Educational Administration and Supervision, 16:135, Feb. 1930.

3/ Marion E. Macdonald, "Students' Opinions as Regards Desirable and Undesirable Qualifications and Practises of Their Teachers in Teacher-Training Institutions", Educational Administration and Supervision, 17:140, Feb. 1931.

Students, judged by the limited number of examples presented, place more emphasis upon the teacher's personal characteristics than they do upon the techniques and mechanics of the teaching process. This does not mean that a method for measuring teacher efficiency should give such items of quality more weight than other criteria. The possibility of such a thing may be considered, but a larger sampling such as would be offered by a random, cross-section sampling of the United States would be necessary if the norms for such criteria were to be valid.

The appended list gives ample proof of the conditions with which an individual would have to contend if he attempted to form a definite opinion as to just what item or quality should be included in a rating scale. The list is taken from the three preceding examples which have only two items out of the total 15 presented that are similar. In other words, 87 per cent of the items are different. If so many differences in opinion appear in these few examples, how many must there be if one was to check up on every method or study made? Standardization of terms is necessary if progress is to be made toward the construction of a valid method. The two similar terms not given below are "Impartial-Impartiality" and "Sympathy-Sympathetic".

1. Skillful in 'Getting ideas across"
2. Human
3. Good disciplinarian
4. Appearance
5. Patience
6. Interest in extra-curricular activities
7. Sense of humor
8. Justice
9. Consideration
10. Friendliness
11. Appropriate dress

Since the writer has commented upon the qualities mentioned by students and has shown their diversity of opinion, he now proposes to follow up the idea in regard to what educators believe about the subject.

Qualities Listed by Educators.--Englehart and Tucker had educators prepare a list of traits which they considered significant for teaching success.[1/] These lists were analyzed and teachers in the Chicago Public Schools were asked to estimate the relative importance of these traits. The final list was used to collect data from 224 students in the high schools of 6 Teachers Colleges. One hundred positive traits were submitted to the students in mimeographed form. They were asked to underline the traits that best fitted their conceptions of what their best and poorest teacher had. Traits which correlated significantly and positively with the quality of teaching according to the traits named were:

1.	Good Judgment	+.93
2.	Clear in Explanation	+.88
3.	Respecting Others Opinions	+.86
4.	Sincere	+.83
5.	Impartial	+.83
6.	Fair	+.82
7.	Appreciation	+.80
8.	Interested in Pupils	+.80
9.	Broad-Minded	+.80
10.	Knows Subject	+.79 °

° 'Strict' was a quality mentioned last on the list of forty-six items and 'Neat in Appearance' was listed forty-second. Mention of the fact is made solely to show that none of the criteria have been standardized so that each item has a definite rank, weight, or value.

1/ Max D. Engelhart and Ledyard R. Tucker, "Traits Related to Good and Poor Teaching", School Review, 44:28-35, Jan. 1936.

At present the assignment of a certain weight to any item is purely an arbitrary and personal matter. As for the above correlations, while they are fairly high, yet, can we accept anything less than +.95 if the results are to be reliable? Hart's findings, however, are in considerable agreement with the above study.1/

Nichols obtained the opinions of 110 superintendents and headmasters of New Hampshire schools as to their criteria for judging good teaching. Six items are given below.

1. Has she a good teaching personality?
2. Community leadership?
3. Health?
4. Skill in teaching technique?
5. Scholarship?
6. General culture?2/

Chittam also obtained from 86 Massachusetts superintendents a composite judgment of what they considered were essential qualifications for a good teacher. A few are given below.

1. Character
2. Personality
3. Training
4. Health
5. Experience
6. Evidence of culture
7. Alertness and enthusiasm

1/ F. W. Hart, "Teachers and Teaching", New York, The MacMillan Company, 1934, pp.131-32; 250-51; 278-79.

2/ A. M. Nichols, "Value and Procedure in Using Schemes for Evaluating Student Teachers with Special Reference to New Hampshire", Master's Theses, Boston, School of Education, Boston Univ. 1932, p.116.

As Chittam mentions in his study, these qualifications are not the sole criteria for judging good teaching for such qualities vary with individual people and, "The finest teachers are those in whom there is a proper balance of desirable qualities".1/

How are we to judge when a teacher has a proper balance of desirable qualities? Upon what valid or reliable basis can we now make such a judgment? Philosophical statements have perhaps as their purpose the awakening of thought upon a subject but something more tangible and definite must follow these abstractions. The implications that may be attached to "proper balance of desirable qualities" could be so diffuse that there would be no limit to the conjectures expressed and no end to the discussion. Apparently educators need to check up on their own opinions and ask themselves, "Just what is it that I actually want to say?"

A few more examples will be offered before the writer comments upon the present confusion of educators in regard to meaning and number of qualities involved in an efficiency method.

Barr lists in his book the characteristic differences of good and poor teachers such as:2/

1/ H. D. Chittam, "The Selection and Appointment of Teachers", Masters Thesis, Boston, School of Education, Boston Univ. 1935, p.84

2/ A. S. Barr, Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies, Public School Publishing Company, Bloomington, Illinois, 1929, p.124.

1. Pupils' interest in subject
2. Physical conditions in room
3. Attitudes of pupils
4. Pupil activity
5. Definite teacher aim
6. Responsiveness of pupil
7. Attitude of teacher
8. General attitude of teacher and pupil
9. Atmosphere of classroom
10. Skill in teaching technique

Armentrout offers still another variation of items.

1. Scholarship
2. Use of English
3. Leadership
4. Instructional skill
5. Psychological method
6. Interest in school life
7. Loyalty
8. Voice
9. Professional growth
10. Initiative 1/

These items had the greatest percentage of agreement between the opinion of training-teachers and superintendents.

Bonser's finding of the qualities desired in teachers by administrators are:2/

1. Good Judgment
2. Self-control
3. Considerateness
4. Enthusiasm
5. Magnetism
6. Adaptability
7. Breadth of interest
8. Honesty
9. Co-operation
10. Refinement

It may be observed that in practically all the examples given in this chapter no mention has been made about judging a teacher through the results of pupil growth or achievement. Only what a teacher is or ought to be, has been used as criteria for measurement.

1/ W. D. Armentrout, "The Rating of Teachers by Training Teachers and Superintendents", Elementary School Journal, 38:511-516, Mar. 1928.

2/ F. G. Bonser, "Qualities Desired in Teachers by Administrators", School and Society, 31:260-262, Feb. 22, 1930.

Apparently those educators whose studies have been submitted believed that the personality traits of a teacher are the only criteria worthwhile for measuring teacher merit. The list below shows how diverse educators' opinions are when judged by the items submitted in the previous studies quoted. The list includes only those items that are dissimilar in meaning.

1. Good judgment
2. Clear in explanation
3. Respecting others opinions
4. Sincere
5. Impartial a/
6. Fair
7. Appreciation
8. Broad-minded
9. Pupils' interest in subject
10. Attitudes of pupils
11. Pupil activity
12. Definite teacher aim
13. Responsiveness of pupil
14. Attitude of teacher
15. General attitude of teacher and pupil
16. Atmosphere of classroom
17. Use of English
18. Psychological method
19. Loyalty
20. Voice
21. Initiative
22. Self-control
23. Character
24. Enthusiasm
25. Magnetism
26. Experience
27. Adaptability
28. Honesty
29. Co-operation
30. Refinement

a/ Impartial and Fair were listed by Englehart and Tucker indicating that the terms have different meanings.

More than 50 per cent of the total number of items used by educators in this list had different connotations. No valid conclusion should be drawn from these figures for the sampling is too small, but the writer believes, for a purpose of general comment, that they do show some indication as to the diversity of opinion.

Knudsen's survey.- Knudesen secured 57 devices from a bibliography of 197 articles on teacher-rating. Information was obtained that:

1. The work was mostly that of professors.
2. The purpose was for a supervisory rather than a promotional function.
3. The method used was obtained through individual judgment and selecting items from other rating-scales.

4. The number and types of people co-operating in its construction were university instructors and principals.
5. There was no evidence of correlation between the ratings of the same teachers by different judges--no validity or reliability.
6. The kind of rating device consisted of scales with their steps defined and the traits classified.[1]

The rank according to frequency of the traits found in the fifty-seven devices were:

PERSONAL EQUIPMENT
Voice, health, etc.

SOCIAL EQUIPMENT
Co-operation, sympathy, etc.

PROFESSIONAL EQUIPMENT
Adequate scholarship
Command of English

CLASS MANAGEMENT
Discipline
Care of physical condition

TECHNIQUE OF TEACHING
Care of Individual, Differences
Skill in assignment, etc.

RESULTS IN PUPILS
Class response
Pupil achievement, etc.

The rank of the various items, all of which have not been given, indicate that the designers of these methods for measuring teacher efficiency placed equal emphasis on teaching and the teacher but not as much on the RESULTS produced in pupils by the teaching process.

1/ Charles W. Knudsen, Evaluation and Improvement of Teaching, Doubleday, Doran and Company, New York, 1932, p.207.

Knudsen says:

> "The final criterion of teaching efficiency should be formulated in terms of the results teachers are able to produce in pupils. Part of these results may be measured, and part, at least for the present, have to be estimated on the basis of the best judgment which can be brought to bear on the matter."1/

Pupil behavior measured by Morrison's group control technique is one objective device for measuring at least one quality of a teacher.

> "There is good correlation', he says,'between the teacher's control technique and his gross effectiveness as a classroom technician....It is fair to assume that a given teacher who has consistently poor control technique cannot be an effective teacher."2/

Nutt favored measuring teacher efficiency through pupil achievement and advocated that definite standards be set up.3/ Any norm would be a standard of some kind of accomplishment and seems necessary in a scientific method, yet Gerling apparently is opposed to any teacher measurement of pupil achievement which is based on an average or norm. His argument is as follows:

> "Teachers should acquire the habit of judging the accomplishment of any pupil, bright or dull, on the basis of individual ability and not by comparison to the work of average pupils or grade norms which are merely mathematical derivatives." 4/

1/ C. W. Knudsen, Op.Cit.p.19.

2/ H. L. Morrison, "The Practice of Teaching", Chicago, University of Chicago Press, 1931, p.134.

3/ H. W. Nutt, "The Supervision of Instruction", Boston, Houghton Mifflin Company, 1930, p.214.

4/ Henry J. Gerling, "Educational Problems", School and Society, 42:104, July 20, 1935.

Perhaps Gerling is right in his belief since the majority of norms obtained are not universal nor nation-wide so that their validity is impaired, but, is not half a loaf better than none? Are not all our judgments based on a standard of comparison? Gerling's point is true if it can definitely be shown just how invalid our educational norms are. Of course a nation-wide sampling for obtaining a norm is the best or most scientific approach to the derivation of a standard.

Jacobs 1/ states in a casual manner that the application of standard tests to students will determine their attainment and supposedly the efficiency of the teacher, but lack of agreement in regard to the details of educational objectives in the high school subjects definitely limit the validity of such achievement tests.2/

Teacher Merit Based on Pupil-Achievement

Student-Achievement Plans.- Some educators contend that standard tests do not reveal the efficiency of a teacher. McCall holds to this view on the basis that conditions for proper testing cannot be obtained. He says, "The actual separation of the contributions of maturity and those of training has never been done." 3/ This statement and the reasoning behind such words would indicate that those men or

1/C.L. Jacobs, Op.Cit.,p. 22.

2/Walter S. Munroe, James C. DeVoss, and Frederick J. Kelly, Educational Tests and Measurements , Houghton Mifflin Company, Boston , 1924,pp.297-331.

3/William A. McCall, Op. Cit. p. 165.

educators who favor the measurements of results by student achievement have no valid or reliable method which will indicate the teacher's efficiency. The National Survey of the Education of Teachers[1] sketches a plan for a desirable study of the measurement of teaching merit in relation to student achievement but it also shows the absence of a reliable measure for many of the variables which enter into the professional success of teachers. The size of the problem is such that they recommend that it be handled by an organization having the time and finances.

Despite these obstacles, some individuals have offered plan, the validity and reliability of which are open to controversy.

Ability Index Plan.--The two main elements of Wetzel's plan are:

1. A comparison of the academic standing of an individual with his ability (Ability index based on reading).
2. Teacher's efficiency determined by dividing average standing of median group into average score of class.[2]

Day based his ratings on student achievement in:[3]

Paragraph meaning
Word meaning
Dictation
Language usage
Literature

This plan is relatively invalid because environmental influences must be present to destroy the true worth of such a measurement. Its use, too, is limited apparently to the elementary school students for the judgment is made solely on a language-reading test.

1/ "Special Survey Studies", U.S. Department of the Interior, Office of Education, Vol.V, Bulletin No. 10, 1933, p.153.

2/ W. A. Wetzel, "Plan of Rating Teachers Based Upon Pupil Accomplishment", School Life, 13:26, October 1927.

3/ L. C. Day, "The Teaching Quotient", Elementary School Journal, 33:606, April 1933.

Nanninga's comment supports the above criticism as:

> "Experiences of various kinds, sometimes not at all associated with classroom procedure or qualities essential for successful teaching, interfere in teacher-rating." 1/

Betts gives five criteria for judging the effectiveness of teaching through the change in pupils as:

1. The subject matter achievement of a pupil-group should increase under the tutelage of a good teacher.
2. The subject matter achievement of pupils, both of the gifted and of the dull, should increase at a rate commensurate with ability.
3. The subject matter achievement of each pupil should increase at a rate commensurate with ability, both in special weaknesses and in special abilities.
4. The extent to which pupils possess socially desirable personal and social traits which promote behavior for the common good should increase under the tutelage of a good teacher.
5. The incidence of nervous behavior should increase in pupil groups under the tutelage of a good teacher. 2/

Item 4 does not seem to be valid or reliable as a criterion for judging a teacher because Hartshorne's study found that inferior home background stood out next to low intelligence as one of the most important factors conditioning the tendency towards deception. 3/ While a student could have a high intelligence score yet his environment could be of such a nature as to counter-act the best efforts of the t teacher.

The conception that data derived from the results of standardized objective tests might be used to determine the changes produced in pupils by the teacher is worthy of consideration.

1/ S.P. Nanninga, "Estimates of Teachers in Service Made by Graduate Students as Compared with Estimates Made by Principals and Assistant Principals", School Review, 36:626, Oct. 1928.

2/ G.L. Betts, Op. Cit. p. 168.

3/ Mark A. May and Hugh Hartshorne, Studies in the Nature of Character , The MacMillan Company , New York, 1928, pp. 283-347.

McCall's Achievement Quotient.--McCall offers a suggestion dealing with the Achievement Quotient based upon reading, writing, and other subjects as well. His formula is as follows: 1/

$$\text{Teacher Efficiency} = \frac{(RQ2-RQ1)\ (Ar.Q2-Ar.Q1)\ldots\text{etc.}}{N}$$

In each case, N equals the number of subjects. For a teacher of a single high school subject the formula would be:

$$\text{Teacher Efficiency} = \frac{AQ2-AQ1}{1}$$

In the above formula, R.Q. is a Reading Quotient; R.Q.1 is the Reading Quotient at the beginning of the teacher's work; R.Q.2 is a similar index after the instruction in reading has been given. The difference between the R.Q.'s is the change in reading ability of the pupils or an indication of teacher efficiency. In a similar fashion Ar.Q. is an Arithmetic Quotient. The teacher's success is the sum of the changes in such quotients and when this sum is divided by the number of subject quotient changes, there is an average of the teacher's general efficiency.

Baird and Bates found that teachers classed as "A" had 16 per cent of their classes in the unsatisfactory accomplishment group; that teachers classed as "C-B" had 61 per cent of their classes in the satisfactory group, and that 22 per cent of the "A" teachers' classes fell below the 61 per cent of the "C-B" group. 2/ This

1/ William A. McCall, Op.Cit.p.153.

2/ James Baird and Guy Bates, "The Basis of Teacher Rating", Educational Administration and Supervision, 15:175-183- March 1929.

shows that the Achievement Quotient technique cannot be taken as a valid proof of a teacher's efficiency.

Crabbs used this type of formula in her study of supervision but the results are not too reassuring.[1] It seems that more experimentation on a wider scope is necessary before there can be any final settlement of the matter. Many variable and uncontrolled factors are present in any achievement quotient plan. Allowances must be made for such phenomena upon a scientific basis before this technique can assume validity. In emphasizing one phase of teaching ability such a conception at least has the germ of consistency about it but this is not the only factor in the measurement of teaching merit. If other factors in a classroom situation involving pupil-teacher relationships could be handled in the same manner through the construction of suitable tests for measuring such qualities, then we would be on the road to the development of a valid, reliable, and objective method. Another factor to be considered, besides the ability to control environmental influences which affect the validity of an achievement quotient is that of a need for more standardized tests to measure student accomplishment at the high school level in ALL subjects. Until such tests are constructed, the Achievement Quotient will have only a limited use and low validity.

1/ Lelah M. Crabbs, "Measuring Efficiency in Supervision and Teaching", New York, Teachers College, Columbia University, Contribution to Education, No. 175, 1925, p.87.

Merit Based Upon Teacher's Exercises

Exercises as Criteria.--Knudsen believes that there is objective evidence of a teacher's ability in the kind of exercises that are used. He suggests classifying them into:

1. Those that furnish a basis for learning activity
2. Those for diagnostic purposes
3. Those for evaluating achievement

He says, "The character of learning exercises is a dominant factor in determining the nature of the learning activities of students and that learning exercises, therefore, have an important influence on the achievement of students."1/

His criteria for evaluating basal learning exercises are:

1. They should be compatible with immediate objectives.
2. " " relate to experiences met in life situations.
3. " " be definite.
4. " " be adapted to individual differences.
5. " " not be too difficult nor too easy.
6. " " be a basis for specific habits, knowledge, attitudes.
7. " " appeal to the pupil and give satisfaction when accomplished.
8. " " be judged on their effectiveness in modifying behavior.

Knudsen's idea seems similar to those educators who favor the "unit assignment" technique for teaching. Both have immediate objectives and the latter has an ultimate one such as that in item eight above. The objective of the unit assignment is the modification of behavior, the acquisition of a skill, or the acquirement of an attitude, concept, or ideal. Such a quality as an "ideal" is very difficult to judge validly or reliably. Although measurement

1/ Op.Cit.p.160.

of a general pattern of conduct is very difficult, for one may understand what is meant by an ideal of honesty but be indifferent as to the acquiring of it, yet Thurstone says that attitudes can be measured with a fair degree of accuracy.[1/]

It is generally held that teachers should have objectives in mind when directing the learning activities of pupils. Morrison states as a principle of operative technique that the teacher should identify the objectives of teaching. This principle is violated constantly in the field of practice for teachers tend to conceive objectives not as abilities to be engendered but as exercises to be done in order to cover the requirements of the course. Knudsen found in an analysis of teacher's exercises using as criteria: (1) Possible learning activities, (2) Possible outcome, and (3) Primary function of exercise, that teachers exploit the use of a type of exercise that calls for a repetition of the symbols of knowledge, that is, they give the answer cue to the pupil; that teachers neglect to provide a sufficient number of exercises that require pupils to generalize their experiences; that they neglect to provide exercises for engendering desirable ideals, attitudes, and concepts by their failure to include in these exercises that which will call out an appropriate emotional response; and that the assignment is given in a manner that neglects to direct the learning activity of pupils with sufficient definiteness.

1/ L. L. Thurstone, "Attitudes Can Be Measured", American Journal of Sociology, 33:529-54, January 1928.

Undoubtedly the study and evaluation of a teacher's exercises is one objective method for measuring a part of a teacher's merit. The diagnostic use of an exercise ought to be considered as well as those which deal with basal learning activity and final achievement. A good reason why the diagnostic concept should be used is shown by the following comment of Judd.

> "Many teachers do not realize the value of an analysis of pupils' errors. They think of a mistake as a purely negative fact; when a mistake occurs, they regard it as their duty to interrupt the pupil's recitation with a view to eliminating immediately the wrong line of thought. It is a fundamental psychological fallacy to think of a pupil's error as a negative fact. Whenever a mistake is made, the pupil has carried on some kind of intellectual process, and that intellectual process is as real as that which takes place when the pupil succeeds in thinking as the teacher wants him to think."1/

The personal qualities of teachers as well as teaching techniques have been presented as a means of measuring teacher efficiency. Some emphasis has been placed by different people upon both criteria yet neither should be over-emphasized; each should have sufficient consideration as a criterion for measuring teacher merit and each should be weighted as to value if they are to be used as an efficiency measure having a scientific basis. What these weights may be must be based on universal norms derived from experimental procedures.

1/ C. H. Judd, Psychology of Secondary Education, Ginn and Company, Boston, 1927, p.135.

Summary

The purpose of this chapter has been to give the reader an over-view of some basic concepts which educators entertain in regard to a method for measuring the efficiency of teachers. The need for standardization of steps and accuracy in the definition of qualities that constitute the criteria for measuring the teacher merit have been emphasized through the comments of Almy, Betts, and Hughes.

Since Da Vinci's time, group judgment has been considered as a means of acquiring a reliable result, but just how many people or what their qualifications for judging should be has not been scientifically ascertained. A recent study of 53 cities by Towner shows that superintendents and their assistants did the judging.

Barr and Emans found in their study of 209 rating scales that there was a great variety of items given by educators and students many of which overlap in meaning because of a lack of scientific definition. These scales also mentioned the personal and social traits of a teacher more than all the other ones.

The writer, by comparing in an arbitrary manner the small sampling of items considered by educators and students as essential qualities for a good teacher, found a diversity of opinion of 87 per cent in the case of students and 50 per cent for educators. The statistical manipulation is neither valid nor reliable but the writer believes it shows some indication of a need for a standardization of terms and also the difficulty at present of anyone attempting to decide what quality or item should or should not be

used in the construction of a measuring instrument.

Knudsen's survey of 57 rating scales showed that the designers placed equal emphasis on teaching and the teacher but not as much on the results produced in pupils. Morrison's group control method as a means of indicating teacher effectiveness has been mentioned and Gerling's opposition to judging pupil-achievement on any average or norm of comparison has been commented upon. The writer has also taken exception to Bett's criteria for judging teacher efficiency through the change in pupils by saying that environmental factors could offset the influence of a good teacher. Other plans such as Wetzel's, based upon an ability index through reading, Day's based upon language-reading ability of students, and McCall's Achievement Quotient based upon dividing the sums of the differences in result between initial and final tests of the subjects studied by the number of subjects, have been presented. The writer has pointed out that McCall's plan or any other plan based upon pupil achievement cannot be considered valid because of the environmental factors which undoubtedly influence pupil growth and development. The findings of Baird and Bates in which 16 per cent of the pupils under good teachers were failing also tend to disprove the validity of the achievement quotient technique.

Knudsen suggests judging teachers by the function of their exercises in regard to their pupils. The criteria for his judgment of such exercises is based upon basal learning, diagnostic, and evaluation of pupil-achievement activities.

While the mention and discussion of the concepts given in this chapter have sought to crystalize thought concerning basic criteria for establishing a method for determining teacher merit, yet nothing definite from a scientific point of view can be expressed as to just what criteria should be used. Experimentation under governmental or other central authority with unlimited finances and with a working personnel of educational specialists is advocated so that nation-wide norms of the qualities to be used as the basis for an objective, valid, and reliable method may be obtained. No individual can produce such a method working alone and no plan, based like McCall's upon pupil-achievement, can have any practical use until standardized tests covering the whole public school field of scholarship are constructed and all the variables except a teacher's ability are controlled.

CHAPTER IV

TYPES OF RATING DEVICES

Graphic Methods

Purpose of chapter.--It is not the purpose here to give an exhaustive list of methods used for measuring teacher efficiency because there are almost as many devices used as there are administrators. Every educator or superintendent must have some method of checking up on his teachers' merit if he is performing his job conscientiously, whether or not his method is based upon a supposedly scientific method of approach or solely on general impression of a subjective nature. No drastic attempt has been made to discriminate between methods as to their good or poor qualities for the examples shown are a random sampling of the writer's reading on the subject, and opinion as to which, if any, method might be the best, can have no scientific substantiation. It is hoped that enough data will be acquired from the small sampling of methods presented to enable the writer to form criteria based principally upon frequency of occurrence. This criteria, while not valid or reliable, will be used as a basis for a general comparison with the current methods now employed in the Massachusetts school systems. The diverse characteristics of these methods necessitate classifying them under three heads as: graphic, numerical-point, and miscellaneous. They will be presented in that order.

A graphic scale.--Bradenburg and Remmers offer a graphic scale having the quality defined on a three-point basis. The list of

Boston University
School of Education
Library

qualities with one illustration are given as follows:

1. Interest in subject
2. Sympathetic attitude toward students
3. Fairness in grading
4. Liberal and progressive attitude
5. Presentation of subject matter
6. Sense of proportion and humor
7. Self-reliance and confidence
8. Personal peculiarities
9. Personal appearance
10. Stimulating intellectual curiosity [1]

Interest in subject

****		
Full of Subject	Mildly Interested	Irksome to him

It will be noticed that no numerical weight or value is assigned. Opinion varies in this respect and a recent thesis of Nichols's advocates that numerical ratings be used instead of a graphic or literal type as illustrated. [2] Nichols' study found that only 4 superintendents put of 78 in the state of New Hampshire used the above type of method. As the reader continues through the chapter it will be noticed that the graphic methods vary as to the number of points on a scale. In the writer's opinion a five-point scale similar to the following example by Brown is better than other types because it should enable the rater to give a more refined judgment. In the field of educational measurements, especially in regard to the

1/ G. C. Brandenburg and H. H. Remmers, "Rating Scales for Instructors", Educational Administration and Supervision, 13:399-406, Sept. 1927.

2/ A. M. Nichols, Op. Cit. p. 139.

grading of pupils, only a system based upon the normal distribution curve is advocated.

A five-point method.--Brown produced a five-point graphic scale. The purpose of such a scale was for the self-improvement of supervisors and was not used as a basis for promotion. It consists of two parts as follows:[1]

1. Personal and Social Qualities
   (A) Personal habits, etc.
2. Methods and Principles
   (A) To what extent have I been successful?

Personal habits

Poor Fair Average Very Good Superior

The following scale by Dragoo is more descriptive in its definition of terms than Brown's but it has only a three-point basis. Such discrepancies in opinion as to how many points should be included on a scale calls attention to the need for standardization.

Descriptive method.--Dragoo made a scale based upon the Purdue and Iowa State College rating form. It includes 17 items but only one is shown here.[2] Evaluation is indicated by a check mark.

1. Preparation .....

Classwork very carefully planned.	Makes plans but lacks ability to arrange properly.	Makes a little or no attempt at preparation.

1/ E. J. Brown, "A Self-Rating Scale for Supervisors", American School Board Journal, 77:36, August 1928.

2/ A. W. Dragoo, "A Rating Scale for Shop Teachers", Industrial Arts and Vocational Education, 21:8, Jan. 1936.

Some concepts of a more interesting nature than any yet described have been advanced by Freyd in the example which follows.

Scoring Stencil.--Freyd drew the following conclusions from his attempt to evaluate outstanding methods of rating:

> "That it is next to impossible to make statistical comparisons between kinds of rating scales..
> .....
> The remaining statistical methods are all subject to errors of interpretation....no figures indicate conclusively that one type of rating scale is superior to another.
> .....
> The comparison of rating scales on non-statistical bases, such as ease of administration and scoring and simplicity of the rating task, is probably as far as we can now go with profit."1/

His own plan deals with 17 traits of personality which are described on a five-point basis of a graphic scale. The traits are:

1. Physique and bearing
2. Neatness in dress
3. Self-possession
4. Sociability
5. Physical energy
6. Alertness
7. Sense of humor
8. Self-assertion
9. Tact
10. Popularity
11. Impartiality
12. Patience
13. Acceptance of criticism
14. Ability to think on feet
15. Fluency of speech
16. Distinctness of speech
17. Interest in teaching

His scale does not concern itself with teaching efficiency as regards technique but only measures the personal qualities of the teacher. An example of the first two traits to be measured follows:

1. How does he impress people by his physique and bearing?

Excites admiration. Very impressive.	Noticeable for good physique and bearing.	Makes satisfactory impression.	Unimpressive, physique and bearing	Arouses repulsion. Looked down on.

1/ Max E. Freyd, "A Graphic Rating Scale for Teachers", Journal of Educational Research, 8:433-439, Dec. 1923.

2. Does he appear slovenly or neat in dress?

Slovenly and unkempt.	Somewhat in- different and careless in dress.	Unnoticeable in dress.	Notice- ably neat and clean.	Fastidious Almost a dude.

The reader will note that Freyd has so arranged his scale that the more desirable extremes of the traits are alternated so as to prevent a rater from checking entirely on one side of the scale. Such a thing could happen when a rater is impressed favorably or unfavorably by the individual who is being judged. Knight found a correlation of+.94 between ratings on quality of Voice and Moral Stamina 1/ and Barr 2/ says, "There is a tendency to skew the rating of every specific trait in the direction of the total reaction of the rater to the subject."

A contrary view is held by Remmers and Brandenburg who drew the conclusion that, "WE find no evidence for the assertion that the halo effect in ratings is enhanced by having the high value of the trait consistently at one end of the scale." 3/ Further experimentation under controlled conditions seems to be the only method to attack and perhaps solve this problem. For purposes of comparison between individuals or between points of improvement of the individual,Freyd has constructed a scoring stencil which gives a numerical expression to the rating . Such a device enables the rater to give a weight or

1/ F.B.Knight, "Qualities Related to Success in Teaching", New York, Teachers College, Columbia University, 1922, Contribution to Education, No. 120, p.63.

2/ A.S. Barr, An Introduction to the Scientific Study of Classroom Supervision, New York, D.Appleton and Company, 1931, p.110.

3/ H.H. Remmers and G.C. Brandenburg, "Experimental Data on the Purdue Rating Scale for Instructors", Educational Administration and Supervision, 13:524, Nov. 1927.

score to his measurement if he is a believer in such a procedure. This stencil is the only one of its kind that the writer has observed in his study. The illustration is given below with the instructions for its use. Ease of scoring results is its best feature.

Freyd's Stencil a/

1 ' 2 ' 3 ' 4 ' 5 ' 6 ' 7 ' 8 ' 9 ' 10

Use this edge for traits 2,3,5,8,10,11,14,15,17.

Scoring stencil
for grahic rating of teachers

Adjust the edge of the stencil to the line on which the rating is made. The score is the figure in the space on the stencil over which the check is made.

Use this edge for traits 1, 4, 6, 7, 8, 12, 13, 16.

1 ' 2 ' 3 ' 4 ' 5 ' 6 ' 7 ' 8 ' 9 ' 10

a/ Reproduction is not full size.

Another method having many more items and apparently more scientific in its approach had been outlined by Mead.

Central tendency plan.- Mead has a type of score card which includes a page of general data about the student-teacher and directions for the use of the card. The evaluation is based upon the normal distribution curve as:

4	3	2	1	0
5% Best teacher	20% Next best teacher	50% Average teacher	20% Next to most inferior	5% Most inferior teacher

The numerical ratings are 4, 3, 2, 1, 0. His complete score card consists of nine main headings each of which are subdivided and described as follows:[1]

I. Result of teacher's work — Ratings
1. Increase in pupil's knowledge (Knowlege equals concepts, precepts, information)
2. Acquisition of desirable habits
3. Growth of pupil's interests
4. Increase in abilities in thinking
5. Growth of ideals
6. Improvement of pupil's taste
7. Use of vernacular
8. Ability to study
9. Co-operation
10. All busy with valuable activity
Central Tendency

II. Teaching Technique
1. Wise choice of aims and related method types
2. Seclection of content
4. Lesson-planning
5. Lesson-assignment
6. Use of texts and other books
7. Skill in use of apparatus
8. Skill in use of method types
a. Drill
b. Problem
c. Project
d. Supervised study
e. Socialized recitation
f. Review
g. Summaries
h. Recitation
i. Appreciation
9. Questioning
10. Adjustments to pupil's differences
11. Variety in method
12. Skill in special method
13. Use of tests
Central Tendency

1/ A. R. Mead, Supervised Student-Teaching, New York, Johnson Publishing Co., 1930, p.481.

III. Scholarship — Rating

1. In subject taught
2. In related subjects
3. In education
4. In general information
5. Use of English

IV. Schoolroom Management

1. Care of heating, lighting, ventilation
2. Care of books and apparatus
3. Seating pupils
4. Keeping records
5. Control of pupils
6. Making records

Central Tendency

V. Professional Interests and Relations

1. Ability in self-criticism
2. Co-operation with co-workers
3. Professional reading
4. Participation in teachers' meetings
5. " " educational associations
6. Upholding of professional ideals
7. Eagerness to improve profession
8. Contributions to publications
9. Advanced study

Central Tendency

VI. Community Relations

1. Cultivation of favorable educational opinion
2. Assistance in community activities
3. Harmonizing with best moral and social standards
4. Discretion in conversation

Central Tendency

VII. Personal Qualities [a/]

1. Sympathy
2. Personal appearance
3. Address
4. Sincerity
5. Optimism
6. Enthusiasm
7. Initiative
8. Fairness
9. Reserve and Dignity
10. Reliability
11. Clear, pleasing voice
12. Industry
13. Adaptability
14. Judgment of moral issues

Central Tendency

VIII. Miscellaneous Factors — Ratings
1. Under teacher's control
2. Not controllable by teachers

IX. Final or Central Tendency of Ratings
a/ All the items are defined by Mead but are omitted here except I-1.

This score card illustrates some essential features that must be included in a method if it is to be serviceable. Such features are: (1) A set of directions, (2) A list of items of quality with sufficient descriptive data to identify them to the one doing the rating. (3) The weighing of items for ease of scoring or statistical manipulation, (4) The use of a standard for comparison such as the man-to-man or best teacher principle.

The validity and reliability of this method is unknown even though Mead makes use of a scientific approach by using a norm or average such as his Central Tendency.

The scale whichfollows is the only example listed in this chapter wherein the author attempted to prove the validity and reliability of his instrument.

Starrak's Device.--He constructed a scale known as the Iowa State Rating Scale. It is given in its entire form below.1/

RATING SCALE FOR TEACHERS

Print instructor's name here---------Department--------Course No.----

Please rate your instructor as to each of the 17 qualities named below. It is necessary in each case only to place a check mark on the accompanying line at that point which seems to you to be right.

1/ J. A. Starrak, "The Construction of a Scale for the Measurement of College Teaching, and the Determination of its Reliability and Validity", Doctor's Dissertation, Boston, School of Education, Boston University, 1932, p.204.

The higher ratings are made by placing the check mark at the LEFT end of the line; the lower, by placing it at the RIGHT. These stages and also the middle ones, are clearly described by the words printed just below the line. The check mark may in any given case be placed either directly above the words or along the line between them as you deem proper. Please make this rating conscientiously and individually. Your instructor will receive only the final summary of the results from the class as a whole. He will use them for self-improvement in teaching.

I. Preparation for class meetings.	Class meetings very carefully planned.	Usually some preparation; often seems inadequate.	Little or no preparation.
II. Interest and enthusiasm in his subject.	Very enthusiastic and interested.	Seems only mildly interested.	Subject seems irksome to him.
III. Ability to arcuse interest in students.	Interest usually runs high.	Students seem only mildly interested.	Majority inattentive most of period.
IV. Organization of course.	Course well organized; parts clearly related.	Some organization but not always clear.	Little or no organization.
V. Scholarship.	Knowledge of subject broad and accurate.	Knowledge apparently deficient at times.	Knowledge very plainly deficient.
VI. Ability to express thought.	Words come easily; meaning always clear.	Some hesitation for words; meaning at times not clear.	Much hesitation for words; meaning often not clear.
VII. Enunciation	Speaks very clearly and distinctly.	Words sometimes indistinct and not easy to hear.	Words very indistinct; often impossible to hear.
VIII. Thinking demanded of students.	Work demands much sound, original thinking.	Thinking and memorization required about equally	Thinking discouraged; much memorization demanded.

IX. Assignments	Assignments clear, reasonable and carefully given.	Rather indefinite and often hurredly given.	Very indefinite usually hurriedly given.
X. Leading discussion and questioning.	Questions thought provoking, discussions lively and worthwhile.	Questions usually call for facts or lead to rambling discussion.	Few questions or none put to class.
XI. Sense of proportion.	Stresses important topics; disregards trivial details.	Occasionally stresses details neglecting important topics.	Often neglects important topics for unimportant details.
XII. Feeling between instructor and students.	Feeling of goodwill prevails strongly.	Neither goodwill nor antagonism seems to prevail.	Instructor tends to antagonize class.
XIII. Sense of humor.	Has keen sense of humor.	Humor occasionally but not often exhibited.	Manifests little or no humor.
XIV. Self-Confidence.	Sure of himself; meets difficulties with poise.	Fairly self-confident; occasionally disconcerted.	Hesitant, timid, uncertain.
XV. Tolerance and Liberality.	Welcomes differences of opinion	Sometimes impatient when students oppose his views.	Easily aroused to temper by opposition.
XVI. Personal Appearance.	Well groomed; clothes neat, clean, in good taste.	Usually rather untidy and careless as to appearance.	Slovenly; clothes and person untidy.
XVII. Personal Peculiarities	Manner pleasing free from annoying mannerisms.	Objectionable mannerisms not serious or numerous.	Constantly exhibits annoying mannerisms.

A reliability coefficient of .86 + .008 was obtained indicating a rathersatisfactory result, but the validity as gauged by the mean difference between the mean scores of students and judges which had a range (mean differences) of 6.58 to 26.06 was not satisfactory.

Starrak says:

> "The data revealed in this investigation would indicate that, in at least a significant proportion of cases, the students' judgment when expressed through the medium of the College Teachers Rating Scale, would not be a valid measure of the teaching ability and personal fitness of an instructor, when the basis of reference is the judgment of experts exercised upon written reports of observed lessons, personality factors being omitted."1/

The above comment shows that the personal characteristics of an instructor do have some influence upon the rating given him by his students, and that the factor must be taken into consideration whenever the validity of a measure is to be determined. It is this inability to control and measure all the variables of a classroom situation which has been one of the obstacles encountered in the effort of educators to construct a method for measuring teacher efficiency which is valid, reliable, and objective.

Numerical-Point Methods

The writer takes the position, in presenting a few examples of methods with numerical values attached to the items or qualities, that the experimentation based on empirical norms from

---

1/ Ibid, p.187.

small samplings or control groups such as the arbitrary point systems to be illustrated, may be advisable for ease of scoring but is not valid nor reliable. Universal norms from random samplings only are justifiable. An arbitrary weight does aid in statistical handling of a measurement but it cannot be used as a reliable indication of the actual merit measured. The three methods which follow illustrate why the writer hesitates to place any confidence in the results of statistical manipulation.

One-hundred point plan.--Ballou's scale for promotion is based upon the following items or criteria each of which has a numerical weight. The total weight equals 100 points.[1/]

1. Scholarship-(20)
   a) Preparation for position held, over and above entrance requirements, and acquired since appointment to present position.
   b) Courses taken in education and in subjects leading to a better preparation for the position held at the time promotion is under consideration.
   c) Contribution to cause of education through articles published.

2. Success in Teaching-(40)
   a) All factors considered in annual rating.

3. Experience-(10)
   a) For credit work must have E rating.

4. Value to School Outside of Classroom-(10)
   a) Work in extra-curricular activities.
   b) Ability in routine record work.
   c) Co-operation with principal and other supervisory officers.

5. Personal Equipment-(10)
   a) Appearance, health, voice, industry, self-control, promptness, punctuality, tact, personality, initiative, adaptability, and judgment.

1/ F. W. Ballou, "Determining Who Are Superior Teachers", Elementary School Journal, 28:256-262, December 1927.

Another plan similar to Ballou's is shown below.

Five-hundred point plan.--Gilmore's device is:[1]

A. Personal Equipment[a]	113 points
B. Social and Professional Equipment	116 "
C. Mechanics of School-Keeping	47 "
D. Technique of Teaching	105 "
E. Results	119 "

a/ There are subsidiary items under these main qualities that are assigned various values.

A good example of the lack of standardization even in the mechanical process of scoring a measure is shown by the two preceding methods. Gilmore assigns a weight of 113 points to the criterion 'Personal Equipment' whereas Ballou gives a weight of 10. It is this lack of knowledge of how much weight ought to be given to a quality and the diverse and arbitrary methods now used, that causes the writer to advocate nation-wide experiments whose results would be in terms of universal norms applicable to all.

Index Card.--Flam shows an index card of qualities rated as:[2]

Personal and Social Qualifications	35
Professional "	40
Technical "	25
Total	100

He draws several interesting conclusions from his study such as:

1/ M. E. Gilmore, "Judging and Rating the Teacher", Educational Review, 74:269-272, December 1927.

2/ August Flam, "Rating Industrial-Arts Teachers", Industrial and Vocational Education, 24:63, February 1935.

1. Score cards for rating teachers have of necessity serious limitations.
2. A perfect score card is a myth.
3. The person doing the rating must be intimately acquainted with the teacher being rated.
4. Many important items upon which a teacher ought to be rated are too elusive to be quantitavely measured.
5. The simplest rating card is, in the last analysis, the most usable one.

The question, "How many items should be included on a measuring device?", may be raised here. Bossing says,[1] "Rating devices with few items to be scored will give a higher reliability than in the case of rating forms with a large number of items." But, we may ask, "How valid would such a device be?" Morris takes the opposite point of view and says, "It is important, therefore, to measure as many significant tendencies of the individual as possible."[2]

This problem is similar to that of giving a weight to a quality because in the writer's opinion only universal norms derived from national experimentation under scientific management and central control will give the criteria needed for solving not only these two problems but all the others involved in the construction of a valid and reliable method. Several methods with such diverse items that they cannot be placed in either of the two preceding groups are listed below.

Miscellaneous Methods

Personal inventory.-Eberhart used a chart on the basis of personal inventory and plan for further growth rather than as a

1/ Op. Cit. p. 126.

2/ E.H.Morris, "Personal Traits and Success in Teaching", New York, Teachers College, Cloumbia University, 1929, Contribution to Education No. 342, p. 13.

method for determining teacher promotion. His plan consists of qualities under six headings. The minor items are not given in detail here. 1/

1. Personal Qualities
2. Social "
3. The Recitation
4. Professional Growth
5. Health
6. In the schoolroom

The following plan is unique in that it is based upon the reactions of the students to their teacher.

Drive and response method.- This plan of Collings is based upon the drive and response of pupils. The purpose was to enable teachers to grow continuously in the process of teaching students. The drive is a measure of teacher stimulation and the response is a measure of teacher direction in the purposeful activities of students. The traits are not shown and almost any interpretation may be given to the criteria listed: 2/

1. Initiation-Evaluation	Choice of Goal
2. " "	Organization of Means
	Execution " "
3. " "	Choice-Consummation of Improvement Leading to Further Goals.

1/ R. Eberhart, "Evaluating Your Own Teaching", Journal of the National Education Association, 20:304, Nov. 1931.

2/ Ellsworth Collings, "A Conduct Scale for the Measurement of Teaching", Journal of Educational Method, 6:97, Nov. 1926.

A lack of definition of terms is apparent in this method and the utility of such a device is limited solely to the owner's use for he alone knows what concepts are implied. A rating device, to be useful, must have reliability. By reliability the writer means that the same results are obtained by different people using the same device.

A statistical method for obtaining reliability is used in the following study.

Correlated study.--Nanninga compared the ratings by nine graduate students of 15 teachers in service with the ratings given by principal and assistant principal, both using the same score card based on:1/

1. Personal fitness
2. Classroom management
3. Teaching skill
4. Achievements
5. General fitness

Nanninga used the Pearson formula and obtained a correlation of +.76 indicating that the findings were fairly reliable, yet the sampling is too small to place much faith in the device. Whether the scale is valid is doubtful, but it is objective, since the personal element (variable factor) of principal and assistant could have influenced their ratings more, so that there would have been no correlation.

1/ S. P. Nanninga, "Estimates of Teachers in Service Made by Graduate Students as Compared with Estimates Made by Principals and Assistant Principals", School Review, 36:622-26., Oct. 1928.

Summary

Every educator seems to have his own favorite theories and ideas concerning the measuring of teacher efficiency. Some go into details and use descriptive terms to define the qualities involved while others allow the rater to exercise his own judgment as to the meaning and implication. Some methods attach a weight or point value to the quality measured while others make use of the graphic type or some other form which the writer has classed as miscellaneous. The purposes of such methods have been for self-improvement of teachers principally but a few use their methods as a basis for promotion.

No evidence is found of actually relating the results to the salary schedule for few educators have confidence in the present methods for measuring teacher ability except possibly their own. Educators do agree on two points: (1) that a reliable and valid method for rating is needed, and (2) that the present methods are not satisfactory.

The purpose of the chapter has been to form criteria from the limited samplings presented based upon frequency of mention. These criteria will be used as a basis for comparison with present Massachusetts conditions as revealed by findings obtained from a questionnaire.

Graphic Methods

Brandenburg and Remmers produced a three-point method with no weight or value given to the result of the measurement except a literal interpretation. Nichols advocates a numerical value but

her opinion is based solely upon the results of her study with the New Hampshire schools where only 4 out of 78 superintendents used letters as a basis for indicating value. Her sampling is too small to indicate any true validity.

Brown has a five-point method for purpose of self-improvement but not for promotion, and Dragoo has a three-point descriptive plan based upon the Purdue and Iowa State rating form. The writer has mentioned the fact that a difference of opinion as to how many points should be used on a scale exists, and he suggests a five-point one because it would give the rater the opportunity to make a more refined judgment.

Freyd has foreseen the need of scoring a measurement numerically and has produced a scoring stencil which enables one to either evaluate teacher efficiency graphically on his five-point scale or else score the results numerically. He believed that rating scales can be compared only on non-statistical bases, such as ease of administration and scoring, and simplicity, and he also attempted to eliminate the halo effect of personal bias by arranging his scale with the desirable extremes of the different qualities alternated. Knight found a correlation of +.94 between voice and moral stamina, and Barr claims ratings of traits are skewed in the direction of the rater's personal reaction to the subject. Brandenburg and Remmers hold to the contrary opinion that the halo effect is not increased by having the high value of the trait consistently at one end of the scale.

It is this controversial attitude and the fact that no reliable criteria or data have been found to direct educational efforts along a central path of accomplishment that is the chief reason why such problems remain unsolved. Individual effort, while essential, is more useful under a co-operative plan of attack.

Mead uses a graphic scale with the evaluation of 4, 3, 2, 1, 0, based upon the normal distribution curve. He uses the mean or average computation for scoring his results and he calls it the Central Tendency. This is one statistical means of arriving at a result and does have the merit of being easy to score, but how valid is it when based upon a comparison scheme as the man-to-man idea? Some norm, of course, is necessary for standardization but only a universal one can be valid.

Starrak's study is the only graphic method listed in which an attempt was made to prove the validity and reliability of the instrument. He found a mean difference in range between the judgments of students and experts of from 6.58 to 26.06, indicating rather conclusively how subtle personal characteristics are and how difficult it is to control the variables in a classroom situation so as to obtain a reliable result. A coefficient of reliability of +.86 $\pm$.008 was found showing that perhaps this method known as the Iowa State rating scale has some use.

Numerical-Point Methods

The writer has taken the position that attaching an arbitrary weight to a quality for purposes of scoring the results is

justifiable perhaps for a general idea of a certain situation but it is not valid or reliable enough to indicate the actual amount of merit measured. Universal norms derived from controlled experimentation is the only method of obtaining a valid weight.

Ballou's plan is based on a 100 point system with the various numerical values attached to the different qualities listed. The purpose of this plan is for the promotion of teachers. The plan of Gilmore's with its 500 point idea shows the lack of standardization in giving a weight to a quality. Gilmore assigns a weight of 113 points to the item 'Personal Equipment' whereas Ballou gives the same item a weight of 10. Gilmore also gives 105 points to 'Technique of Teaching' while Flam in his index card gives 25 points. These are only a few reasons or examples why the writer favors a policy of national, controlled experimentation. Flam's comment upon a simple score card has raised the question, "How many items should be included on rating scale?" Flam and Dossing advocate few items whereas Morris says as many qualities as possible should be used. There have been no scientific findings either as to the number or kinds of items to be used but the number under present conditions would seem to be reached when the device becomes too cumbersome to operate.

Miscellaneous Methods

Several methods of a miscellaneous type have been presented such as Eberhart's and Collings's which had for their purpose teacher improvement and growth. Colling's criterion is so indefinite

that only he has a concept of just what the terms should or ought to mean. The results of his plan are based upon the drive and response of the students,--the drive being a measure of teacher stimulation and the response a measure of teacher direction in the purposeful activities of pupils. Too many uncontrollable variables influence such findings, such as: mental attitude of teachers and students, intelligence of students, room conditions, the appeal of subject alone to the student, and many others.

Manninga obtained a reliability coeffici nt of +.76 between the ratings given by students and those by principals using the Pearson formula, but his sampling of only 15 teachers is too small to have any validity.

One might form a basis for the construction of a method by adopting those qualities which appear in the order of frequency on the various measuring devices. Educators have done this before and have not obtained any worthwhile results. Standardization is needed and must be secured through some progressive plan of experimentation based on nation-wide samplings of criteria.

The following criteria based upon frequency of mention have been chosen arbitrarily by the writer without any valid or reliable foundation for such action except to make a general comparison with the items on the rating devices used at present in some of the Massachusetts school systems. They are:

Do the Massachusetts methods use such criteria as:

1. Personal Qualities of Teachers
2. Teaching Techniques of Teachers
3. Results of Teaching
4. Professional Interests of Teachers
5. Community " " "
6. School Management " "
7. Do they weigh their items?
8. Do they use a graphic method?
9. Are the results used as a basis for promotion? Supervision?

CHAPTER V

MEASURING TEACHER MERIT IN MASSACHUSETTS

The Questionnaire

Scope of questionnaire.--In order to determine what the current practices were of the Massachusetts school systems in regard to the measuring of teacher efficiency, a questionnaire was sent to 100 superintendents. The three principal cities of the state, namely, Boston, Worcester, and Springfield, were not included in this survey because the writer believed that if these cities did have a method for determining teacher merit, the findings would not be typical of the state as a whole. The location of these towns and cities was based upon the geographical factor as well as the 1930 Federal census so that there would be as wide a sampling as possible with the limited resources available. The survey covered the following communities, the location of which are shown on Page 74-a:

Group 1	20 cities	
" 2	42 towns with a population over 5,000	
" 3	38 " " " " under "	but maintaining high schools.

Limitations of study.--Since there were 203 superintendents in service in Massachusetts before the present gubernatorial regime was established, the small sampling of 76 replies from 100 questionnaires sent out by the writer will have no determined validity. No statistical method of comparison is used for the writer believes there must be a valid and reliable collection of facts before any statistical treatment can be used that would be

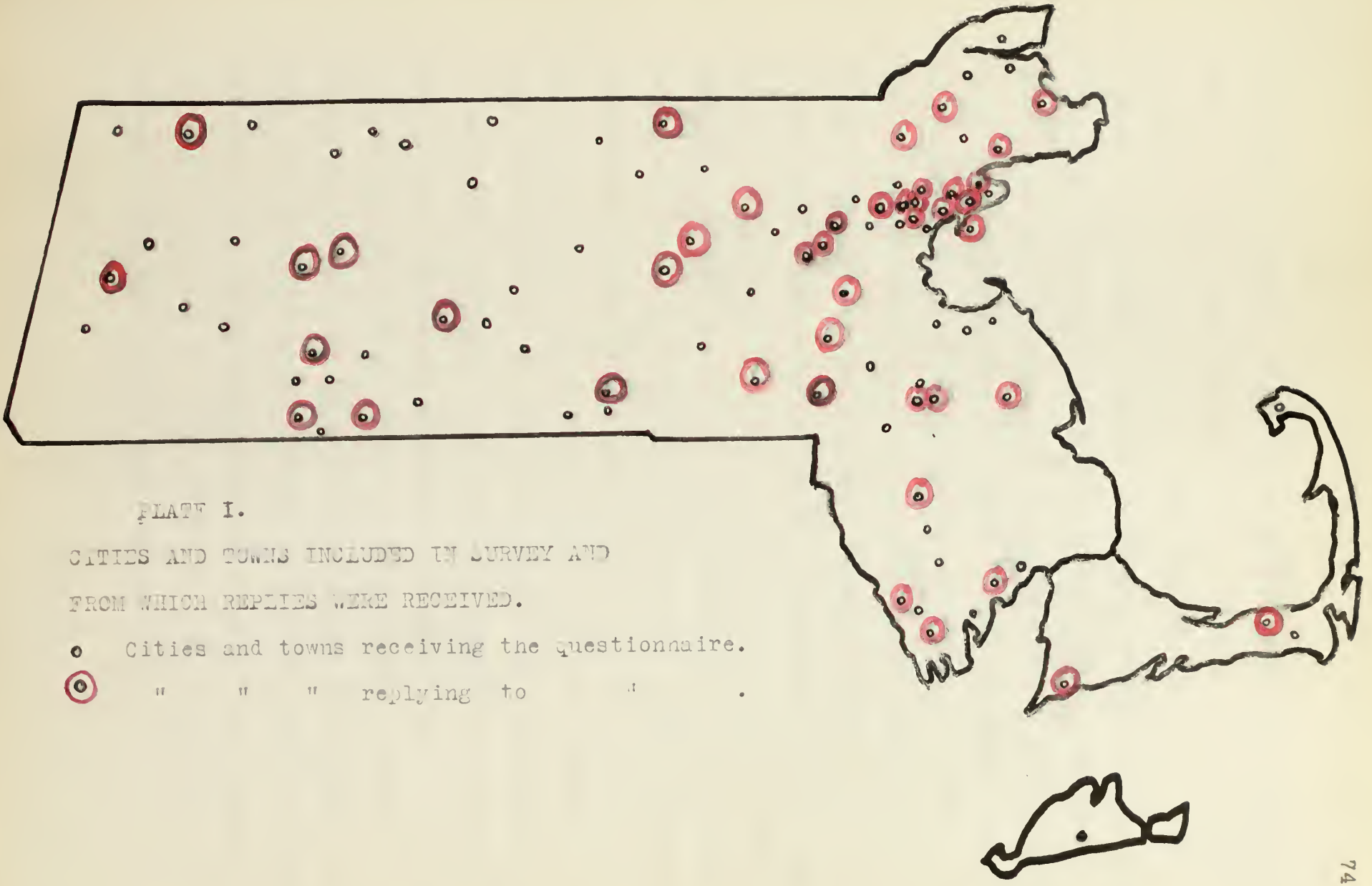

PLATE I.

CITIES AND TOWNS INCLUDED IN SURVEY AND FROM WHICH REPLIES WERE RECEIVED.

o Cities and towns receiving the questionnaire.

◎ " " " replying to " .

worthwhile. Definite objective evidence of a valid and reliable method for measuring teacher merit is lacking. Even definite knowledge as to what qualities should be included in a device is lacking, for personal opinion, subjective evidence, and studies using only small samplings are available. That technique or practice is best theoretically that improves the teaching efficiency of a school system but before such a technique can be created, definite understanding of the constituents of good teaching must be ascertained. There is no statistical substantiation of the best opinions on the subject. Because the majority of rating devices list the quality 'tact' is no reason to believe that that quality is one valid or reliable attribute of a good teacher. Then, too, the replies of some superintendents do not indicate that their practices follow their theories. Some superintendents, in answering the questionnaire, claimed to have a method for measuring teachers based on merit but only a few submitted evidence upon which to justify their claim. The questionnaire data is unreliable at least for statistical treatment for varying interpretations of questions, dissimilar opinions, and the professional ability and attitude of educators all affect the results. The writer merely attempts to sift what opinions are stated or held about the problem both in previous literature and the replies received from the questionnaire, and then form some general conclusions. At best, the conclusions will be illustrative of present trends and thought upon the subject.

Source of data.-- TABLE 1 shows how the 76 superintendents replied to the questionnaire.

Table 1. Proportions of Superintendents Employing Various Devices in the Selection and Promotion of Teachers.

Practices Employed in the Selection and Promotion of Teachers.	Superintendents Reporting Practice	
	Number	Per cent
Systematic, objective method of determining probable teaching ability of candidate	17	20
Systematic, objective method of determining the merit of teachers in service	12	16
Materials sent explaining methods	12	16
Salary schedule based on teacher merit	7	9
Copy of schedule and criteria on which salary increases are based	20	26
Summary of results of this study	54	71

Only 20 per cent of the superintendents who replied to the questionnaire believed that they had a systematic and objective method for judging the probable teaching ability of candidates, and only 16 per cent believed that they had such a method for their in-service teachers. When asked to supply material or printed matter explaining their methods, only 16 per cent responded.

The third question, "Have you in actual operation a salary schedule based on teacher merit?" could be ambiguous in that the superintendent could have a salary schedule based on merit but it would not be complying with the criteria mentioned in the previous chapter. For instance, the superintendent might base his schedule on the professional improvement of a teacher yet this criterion is only one of the many items to be considered in measuring a teacher's efficiency. Few superintendents apparently applied this interpretation to the question for only 9 per cent acknowledged that they had a salary schedule based on merit.

Twenty superintendents sent copies of their schedules and 12 of them submitted criteria upon which the increases in salary were based. It is apparent from the above limited sampling that few superintendents have any systematic or objective method for determining either the ability of beginning teachers or those already in the service; nor have they a salary schedule based on merit.

That the questionnaire dealt with a problem of some interest is evidenced by the 54 superintendents who desired a summary of the findings of this study.

Teacher turnover in the United States in 1931 was about 12 per cent which means that from 100,000 to 120,000 new teachers are needed each year.[1/] This need shows that some systematic method for selecting these new teachers ought to be used. Perhaps

1/ The Status of Teacher Demand and Supply in the Elementary and Secondary Schools, Research Bulletin of the National Education Association 9:338-349, Nov. 1931.

the turnover figure is not so large since 1931 for economic conditions may have stabilized the profession more than in the past. If this stabilization is an actual fact, then all the more emphasis should be placed upon the need of superintendents for having a method for measuring the merit of their in-service teachers. Big business interests do not tolerate inefficiency and surely our school systems should not when the sum involved in education amounted in 1930 to 2,321,728,645 dollars.[1] The taxpayer, citizen, and parent ought to receive the best returns from their money but will they if no method for weeding out the inefficient teacher is devised?

Although Table 1 shows only 9 per cent of the superintendents replying who have a salary schedule based on merit, yet the importance and the benefit to be derived from such a schedule, if it were valid, is incalculable. A schedule based as a result of a valid, reliable, and objective measure of teacher merit MIGht eliminate those teachers not professionally minded and would encourage the successful teachers to further effort. If the schedules were similar to that of the Providence plan that will be presented in chapter six, then there would be no limit to the salary increments for successful teachers except their natural span of life.

1/ State Support for Public Education, The National Survey of School Finance, U.S. Office of Education, Washington, D. C., 1933, p.309.

The political influence would be much lessened, too, for the efficiency of the candidate or teacher in-service could be definitely presented by the superintendent to his school committee without fear of equivocation.

Table 2 shows the number of superintendents who sent back usable data.

Table 2. Towns and Cities from which Usable Replies were Received.

Group	Number in Group	Number Replying	Per cent Replying
Cities	20	12	60
Towns over 5000 in population	42	18	43
Towns under 5000 in population having high schools	38	16	42
Total	100	46	46

The writer's belief that the findings would not be typical of the state as a whole if the larger cities were included in the study seems justified by Table 2 in which 60 per cent of the cities responded with data whereas only 42-43 per cent of the smaller communities furnished any information. The validity of such a small sampling is questionable and the writer does not place too much confidence or attempt to read into the data too many interpretations. It is hoped that the data may give some information that is interesting even if not scientifically useful.

Criteria for judging candidates.--Table 3 shows the criteria used by 17 superintendents for judging prospective candidates. Not all of the superintendents who claimed to have a method gave the complete evidence to prove their contention.

Table 3. Criteria Used by 17 Superintendents in Judging Prospective Candidates

Criteria	Superintendents Employing this criteria	
	Number	Per cent
Training School Record	12	70
Experience	11	64
References	11	64
Interview or Observation, or both	7	41
Written Examination	3	17

It is evident from Table 3 that some superintendents place more weight or value upon training, school record, experience, and references than upon either the interview, observation, or written examination technique. The Research Bulletin of the National Education Association for 1932 found that only 3 per cent of the cities in the United States required written examinations for candidates and that only the larger cities demanded it.[1/]

1/ Administrative Practices Affecting Classroom Teachers: The Selecting and Appointment of Teachers, Research Bulletin of the National Education Association, 10:24-27, January 1932.

The writer's study found only three superintendents in Massachusetts using the written examination and two of them were from the larger school systems. Chittam also found the same number as the writer did, the only exception being that the three cities using such methods in his study had a population of over 100,000.[1] Apparently educators are satisfied that if a candidate has been graduated from an approved normal school or college no examination is necessary.

An interesting example of the value of examinations is shown in Table 4. The results of several correlation studies of the relationship between test scores and subsequent ratings of teacher efficiency were gathered together by Butsch.[2] The uncontrolled variable is, of course, the rating of teacher success hence such findings need to be cautiously interpreted.

Table 4. Correlation between Examination Scores and Teaching Success.

Investigator	Coefficient of Correlation
Knight	+.61
Knight and Ruch	+.38
Hunt	+.30
Boardman	+.28
Tiegs	+.02

It is observable that not only are the correlations between tests and success as a teacher low [redacted], perhaps

1/ M. D. Chittam, Op.Cit.p.148.

2/ R. L. G. Butsch, "Teacher Rating", Review of Educational Research, 1:99-107, 1931.

proving that tests are invalid as a criteria for judging teacher merit, but also that Knight, who found the highest correlation +.61, and Tiegs the lowest +.02 both used the same tests in their studies; the Knight, Bathhurst, Ruch, and Telford Standardized Test for Elementary Teachers, University of Iowa. Despite these findings it is not safe to entirely condemn examinations for as tests become better standardized with higher validity and reliability, it is possible that they may be of great value in the selection of teachers.

Methods of selecting teachers.--Although this study is not interested primarily in the selection of beginning teachers yet a few examples and comments of various superintendents will not be out of place here. The majority of superintendents who contributed some ideas on the subject based their judgments upon such criteria as: school grades, professional training, interviews, personality, references, and ability to take charge or participate in extra-curricular activities. A few examples are given below.

One superintendent writes:

> "The probable ability of candidates for positions in our schools is determined through careful investigation of their qualifications. In the case of candidates without teaching experience copies of their college records are almost invariably obtained, estimates of their ability by college instructors and others acquainted with their work is sought and personal interviews are arranged for prior to appointment."

Another says:

> "We investigate with great care all candidates--and base our selection on the school and college record, their previous experience and record, their character references, their personality and appearance."

One comments tersely, "Standard tests in English and professional knowledge."

Although measuring the merit of teachers has been discussed in the earlier part of the chapter and reference been made on page 78 to the sum spent in education in 1930, yet in Massachusetts only 12 superintendents in answering the questionnaire claimed to have any criteria for judging their in-service teachers.

Rating teachers in service.--Table 5 shows the criteria upon which the 12 superintendents based their judgment of teachers.

Table 5. Criteria Employed by Superintendents in Judging Merit of Teachers in-Service.

Criteria a/	Superintendents Using Criteria	
	Number	Per cent
Professional Study	7	58
Observation	7	58
Rating Scales	5	41
Progress of Pupils	1	8
Experience	1	8

a/ One superintendent gave the Strayer-Engelhardt form published by the C. F. Williams and Sons Company, Albany, New York.

Although the limited sampling submitted is not a valid finding for the state as a whole, nevertheless it shows that some superintendents place more emphasis or value upon professional study, observation, and rating devices than upon pupil achievement and experience. The writer believes that if a rating method is

produced that is both valid and reliable the ultimate trend will be towards using it as a basis for all administrative purposes dealing with promotion, dismissal, and salary of teachers. The reason why educators have turned to professional study as a criterion for judging teacher merit is because no method has yet been devised that is valid and reliable.

It may be fair to assume, since only one superintendent mentioned experience as a criteria for measuring teachers in-service, that it is not always of a beneficial character and is not necessarily an indication of a good teacher. The use of pupil achievement plans discussed in previous chapters was suggested by only one superintendent. Perhaps superintendents included this factor under observation but whether observation means looking for good teaching technique, response of pupils, classroom management, or any other quality connected with the teaching act must be surmised. A clear understanding by definition of terms is needed.

Some superintendents, in commenting upon the problem of measuring efficiency, brought up the question of the appointment of local teachers. Their remarks show only too well how the economic and political factors have influenced administrative procedures.

One says: "We have all local teachers and could not make a distinction as to merit."

Another comments, "Why talk about the selection of teachers when so many local candidates are given positions to teach in the schools of Massachusetts."

Still another says, "At the present time, most superintendents find that the local candidate has to be given preference, for many committees put the individual before the cause."

A good example of the feeling towards non-resident teachers is illustrated by the following article which appeared recently in a Boston newspaper.[1]

> "I am not trying to criticize the ------City council or the ------city officials for desiring to discharge a non-resident of that city working in their school department. I am of the opinion that they are absolutely correct and I certainly trust that our school committee in -----will take similar action in not only discharging residents of ------but every non-resident coming from other cities and towns."

If ever a need for a valid and reliable method for measuring a teacher's ability is felt it is at the present time under the circumstances illustrated above, where politicians and not educators are running our schools.

A good illustration of the subjectivity of present methods used by superintendents is given by the following quotation:

> "I really have no systematic and objective method for determining teaching ability or success. Although I am open-minded on suggestions for such a method, to date I have not found it necessary nor particularly helpful.
>
> Having once been satisfied with the character, scholarship, personal appearance and personality of the teacher, I then must get rather a general impression of her ability in the classroom by observing her teach. I have never endeavored to analyze particularly what makes up my impression because I have not found it necessary.
>
> Some of the most important things which make up the impression however, are obviously the teacher's understanding of those whom she is teaching, a cheerfulness, and genuine interest

1/ The Boston Globe, Saturday, March 7, 1936, p.13.

> in her work. The rapport or relationship between herself and her students is a very important factor and can very quickly be determined, at least by an experienced observer of classroom work."

The previous statement points to the fact that a 'general impression' of a teacher seems to be a conglomerate mixture of subjective attitudes formed neither under controlled conditions nor by scientific analysis. However, it must be admitted that certain administrators may be unusually successful in estimating teacher success by such subjective methods. It is this individual laissez-faire policy having no status upon which to base its judgments that accounts for many suggestive but no practical, definite plan for attacking this problem of measuring validly and reliably a teacher's ability.

Salary schedules.--One of the first facts to be observed from the questionnaire was that the salary schedules did not exist or that they were reduced. This was especially true in towns where the effect of the economic depression has been such that no salary schedule could be maintained. Such comments as, "No such animal since 1931", were often made by the superintendents of these smaller towns. The other larger communities, both town and city, have been forced to reduce their salary schedules anywhere from 5 to 20 per cent. Some localities at the present time are replacing these "cuts", but nothing definite can be ascertained by this study because the majority of the salary schedules sent by the superintendents are representative of a period from 1929-1934.

It is evident from the limited scope of this study that salary schedules are principally based upon professional study and years of service. Automatic increases based on "General satisfaction" in the teacher appears to be the most common method employed. Twenty-three of the superintendents, or 33 per cent, who supplied any information about their salary schedules mentioned the above criteria. These annual increments ranged from 40-100 dollars with the 100 dollar increment being the most common.

Table 6 shows the criteria used by 7 superintendents who believed they had a schedule based on teacher merit.

Table 6. Criteria on the Basis of which Salaries are Increased, as Reported by 7 Superintendents

Criteria	Number of Superintendents
Observation General Satisfaction to sup't and committee.	4
Professional Study Successful Experience.	4

The overlapping in meaning is apparent in Table 6 where 7 superintendents have given 8 replies. The surmise, while not valid may be considered, that perhaps observation and successful experience are related to each other when judgment is made by some educators, and professional study and general satisfaction mean something else to others.

Professional study may be an adequate proof of a teacher's interest in her work and it may also be a mechanical gesture to secure a salary increase. As a criteria for salary increases it cannot be valid nor reliable until some method is produced that can validly show, through measurement, that the studies pursued have benefited the teacher.

The general character and looseness of meaning implied by the expression "observation and general satisfaction to superintendent and committee" involves so many, many, uncontrolled variables that affect both the individual being judged and the group doing the judging as to render the result of such findings invalid and unreliable.

Group judgment does act as a check to the biased opinion of an individual, and some Massachusetts superintendents recognize this fact for 11 out of 19 administrators who observe their teachers make use of two or more judges. Educational theory today is that the school principal should devote approximately 50 per cent of his time to the supervision of his teachers. A few quotations from superintendents will show the general attitude expressed in regard to who should do the supervising.

> One says,"No teacher with experience is engaged unless superintendent, principal, or supervisor has had an opportunity to observe her work in the classroom. We do not have a rating scale. Salary increments in salary schedule depend upon completion of professional improvement work as well as on satisfactory performance of classroom work."

Another writes,"I would say that our only method is that of supervision by principals and supervisors."

One states, "After becoming members of our staff, they are subject to the scrutiny of our supervising principals and of myself."

Another comments, "Teachers are promoted according to the recommendation of the superintendent. I have 60 teachers under my supervision. I visit them in their classrooms very often. Their success in the classroom determines their promotion."

Another replies, "Full time principals observe substitute teachers and supervise regular teachers, turning in reports regularly."

Although the above comments do not indicate by what means the supervision in detail was made, still the writer received six methods used by superintendents which will be presented. The sampling is so very small that it cannot be taken as typical of present Massachusetts practices.

Representative Methods for Rating Teachers

Falmouth.--A three-point rating scale for supervisor's estimate of a teacher is used. The items are undefined and of a general nature allowing the rater to read into the terms his own subjective interpretations. It is as follows:

	Satisfactory	Doubtful	Unsatisfactory
Preparation			
Presentation			
Pupil Reaction			
Management			
Co-operation			
Discipline			
Progress			

If a comparison is made with the criteria listed on page 7[illegible] of this study, the reader will see that; theabove method does not have many personal items or qualities, nor does it have items dealing with professional and community interests; the items are not weighed; the method is not a graphic one and the purpose is for supervision only. It does, however, emphasize the teaching technique and the results of teaching, and also school management.

Andover.--A five-point method with each item defined and classed under five main headings with superintendents and principal doing the rating are the essential characteristics of this plan.

DETAILED RATING	Poor	Fair	Good	Very Good	Superior
I. Personal Qualities					
1. Presence					
2. Enthusiasm					
3. Initiative					
4. Industry					
5. Resourcefulness					
II. Social and Professional Equipment					
1. Co-operation and Loyalty					
2. Daily Preparation					
3. Mastery of Subject Matter					
4. Use of English					
5. Professional Interest and Growth					
III. School Management					
1. Care of Health Conditions					
2. Neatness of Room					
3. Control of Room					
4. Economy of Time and Effort					
IV. Conduct of Recitation					
1. Skill in Motivating					
2. Teaching How to Study					
3. Stimulation of Thought					
4. Organization of Subject Matter					
5. Drill Effectiveness					
6. Attention to Individual Needs					
7. Lesson Assignment					
V. Results					
1. Moral Influence of Teacher					
2. Attention and Response of Class					
3. Growth of Pupils in SubjectMatter					
4. General Development of Pupils a/					

a/ All the items are defined but only one is given here. "General Development" means--Increase in pupil's ability and power along lines other than those of subject matter.

This card contains everything listed in the criterion for the comparison except that it does not weigh the items, is not a graphic method, and does not mention community interests as a criterion for judging the teacher. In comparing the Andover method with the preceding one it is noticeable that opinion differs as to how many items should be included in a scale. Andover has 25 items and Falmouth only 7, showing that the need for standardization is true in Massachusetts as well as over the entire country.

Rockland.--A four-point plan with a list of 16 items is given as follows:

1. Character and Conduct
2. Scholarship
3. Personality
4. Ability to get along with other people
    a. Other teachers
    b. Parents
    c. Pupils
5. Tact and sympathetic understanding
6. Culture and Refinement
7. Self-control and poise
8. Loyalty and co-operation
9. Ability to discipline
    a. Class room
    b. Study Hall
    c. Home Room
    d. Corridor
10. Teaching skill--attention to individual needs
11. Ability to get pupils to prepare work regularly
12. Ability to get pupils to make up work
13. Attendance at school functions
14. Punctuality
    a. At school
    b. In handling reports
    c. In keeping plan book
15. How does he accept criticism?
16. Health and vitality

RATING: Excellent, Good, Fair, Poor.

The criteria listed emphasizes the personal qualities of a teacher more than any of the other items used as a basis for comparison. No weight is given these criteria, the method is not a graphic one, and the purpose is only for a supervisory function.

Melrose.--The following rating card has a four-point definition of qualities with the rating based upon the composite value of the whole list of criteria. The rating card has 24 items covering personal qualities, teaching technique, and school management with more emphasis placed upon the personal attributes of the teacher. The items are given no weight, and the method is not graphic but it does base promotion upon the results of such ratings.

	S	G	F	P
1) The Teacher Herself				
1. Appearance				
2. Voice				
3. Health				
4. Professional Growth				
5. Use of good manners				
6. Promptness, accuracy, dependability of reports				
7. Co-operation				
8. Attitude and response to criticism				
9. Enthusiasm and Industry				
10. Tact, courange and judgment in handling difficult situations with pupils and with parents				
11. Personality as a whole				
2) The Classroom				
1. Attention to comfort and health of class				
2. Evidence of good housekeeping-				
3. Evidence of work in progress--				

S G F P

3) The Class
   1. Freedom from repression and formality ----------
   2. Interest and attention --------
   3. Spirit of co-operation --------
   4. Growth in self-reliance -------
   5. Growth in habits of courtesy, good health, citizenship ---

4) The Teaching
   1. Preparation-regularly done -----
   2. Use of effective and modern methods ----------
   3. Adapted to needs of individuals
   4. Regularly checked by tests ----
   5. High in standard ----------

In light of the following definitions, is the teacher on the whole rated by you as Superior _____ Good ____ Fair ____ Poor ____.

Superior = personal qualities, professional attitude, skill in teaching, and ability to get from her pupils favorable reactions to school requirements must all be of a high order.
Good = fully satisfactory
Fair = moderately satisfactory
Poor = unsatisfactory.

Longmeadow.--The following service record has only six main criteria which are subdivided by definition into 19 qualities. A five-point method for evaluating the criteria is used and salary increments are granted only in case the items can be honestly rated as satisfactory and that there is an earnest endeavor to make improvement each succeeding year. The personal attributes of a teacher are not emphasized in this method as they have been in the preceding ones. The teaching act including management, technique, and results appear more important than any of the other criteria. The items are not weighted; the method is not a graphic one; and the results are used as a basis for promotion.

E G F P U

1. Satisfactory Classroom Management
   a. Organization of the work -------
   b. Organization of the children for their work ----------
   c. Orderly procedure of the recitation ----------
   d. Orderly procedure of the study groups ----------
   e. Approved teaching techniques ---
   f. Personal influence with pupils--

2. Plan Book:
   a. Work planned one week in advance
   b. To show amount and methods used-
   c. The year's work complete at end of year ----------

3. Satisfactory Management of the Following:
   a. Books, supplies, work materials-
   b. Neatness of room, desks, tables, blackboards ----------
   c. Registers and records ----------
   d. Punctuality and stimulating the ideal in children ----------
   e. Co-operation with parents, teachers, supervisors and school officials ----------

4. Professional Spirit as shown by:
   a. Search for latest improved methods ----------
   b. Practice of improved methods ---

5. Results of Teaching:
   a. Pupils' present achievement ----
   b. Permanency of results ----------

6. Health

E = Excellent P = Poor
G = Good U = Unsatisfactory
F = Fair

North Adams.--The superintendent of this school system submitted the only graphic rating method. It is a five-point scale consisting of 12 items with the items briefly defined and weighted.

The criteria deal almost entirely with the personal qualities of a teacher.

Rating Scale

Each of the items listed is rated on a five point scale on which 1, 3 and 5 have been briefly defined. Please place an "x" directly beneath the number which best represents an accurate judgment.

	1	2	3	4	5
1. Health	Vigorous-Never Absent		Average-Rarely absent		Frail-Often Absent
2. Character	Highest ideals and conduct		Conforms		Weak-Unstable
3. Personality	Impressive-Stable-Attractive		Ordinary		Repellent Nervous Disliked
4. Force	Very marked-Makes things move		Average Sufficient		Helpless Follows does not Lead
5. Co-operation	Genuine-Intelligent-Effective		Fair-Reciprocates in Ordinary Routine		Lacking. Unwilling. Trouble Maker
6. Tact	Meets difficult situations frankly without Offense		Meets ordinary situations Satisfactorily		Cannot get along. Many Antagonisms
7. Professional Knowledge	Exceptional-Knows Newest and Best		Average-Knows Generally Accepted Things		Deficient-No Evidence of Study or Growth
8. Mental Power	Brilliant		Average		Stupid-Dull

	1	2	3	4	5
9. Teaching Type	Leads-Inspires Motivates		Ordinary-Combines Motivation and Compulsion		Compulsion Type
10. Teaching Ability	Expert.Resourceful Maximum Results		Average.Ordinary Results		Lacking-Pupils Show Little Gain
11. Culture	Marked Interest in and High Appreciation of Finer Things		Ordinary		Low Interest and Tastes
12. Capacity for Growth	Very Great		Little-Needs Prodding		At Limit of Growth

This method with its emphasis upon a teacher's personal attributes calls attention to the fact that three out of the six methods have attached more value to this criterion than to the others. In a previous part of the study the writer has advocated that no quality or group of qualities should overbalance or outweigh other groups. It may be the wrong attitude to take for the evidence here indicates that educators place more emphasis on these personal qualities. However, until universal samplings of proven validity are submitted, the writer will be loath to change his viewpoint.

Another problem raised previously and again rising here is that concerning the number of points on a scale. The writer's attitude that a five-point scale enables the rater to give a more refined judgment seems to be supported here, for three out of the six methods used it. The sampling is too small and invalid for the

writer to feel confident that this attitude is correct but it will be maintained until some scientific findings prove it to be false.

The question as to how many items should be included in a scale apparently varies with the concept of the individual constructing it. The fact that the number of items in the six methods varied from 7 to 25, shows that diversity of opinion exists in Massachusetts the same as in other localities and serves to emphasize the writer's thesis that experimentation on a national basis can be the only way for solving the problem.

The following items gathered from the six methods give ample proof of a need for definition and standardization. The qualities of health, co-operation, and daily preparation had the greatest frequency of mention.

Presence
Enthusiasm
Initiative
Industry
Resourcefulness
Professional Interest
Care of Health conditions
Neatness of room
Control of room
Economy of time and effort

Attention to individual needs
Lesson Assignment
Moral influence of teacher
Attention and response of class
General development of pupils

Force
Growth of pupils in subject matter
Tact
Training
Reliability

Use of English
Mastery of Subject Matter
Daily Preparation
Loyalty
Co-operation
Skill in motivation
Teaching how to study
Stimulation of thought
Organization of subject matter
Drill effectiveness

Presentation
Management
Discipline
Progress
Health

Voice
Personality
Scholarship
Mental power
General intelligence

Dress
Ability to get along with other people
Sympathetic understanding
Culture and refinement
Self control and poise

Teaching type
Teaching ability
Use of good manners
Appearance
Neatness

Capacity for growth
Promptness, accuracy, dependability with reports
Attitude and response to criticism
Courage and judgment in handling difficult situations with pupils and parents

Evidence of work in progress
Results of teaching:
Permanency of results
Ability to get pupils to prepare work regularly
Ability to get pupils to make up work

Attendance at school functions
Organization of children for their work
Orderly procedure of the recitation
Orderly procedure of the study groups

Personal influence with pupils
Class: Freedom from repression and formality
Interest and attention

Use of effective and modern methods
Regularly checked by tests
High in standard

Spirit of co-operation
Growth in self-reliance
" " habits of courtesy, good health, citizenship

The list of 72 items gathered from the six methods of Massachusetts superintendents shows only too well how vague, diverse, and indefinite thought has been upon the subject.

Although some educators do place confidence in rating schemes for promotional and supervisory practices, nevertheless other superintendents desire to have no rating plan associated with their judgment of teachers.

One says: "Thirty years of experience with a low percentage of failures" is reason enough for his not using one.

Another replies, "I believe that this can never be measured objectively. Teaching school is so closely connected with

personality factors that I find subjective judgment more accurate."

Another comments, "We have considered seriously putting into effect a salary schedule based on merit, but have NOT done so in the belief that it would not work. It sounds good, but has a demoralizing effect on teacher morale."

One superintendent's remark offers a challenge to one interested in the construction of a valid, reliable and objective method and shows that the disdain held for present methods would cease if a definite achievement in research could be made. He says, "I consider any objective method I have seen worthless. Would like a good one."

Another, in speaking about a method for candidate merit, "Maybe not but I am quite willing that the merit of teachers selected and in service should speak for effectiveness of methods of selection."

The last superintendent quoted says, "Have not found any satisfactory method. Rely on my own personal judgment."

As the situation now stands, educators recognize the need for a valid and reliable method and are dissatisfied with the ones now being used, and only a few use them as a basis for promotion or salary increment.

Summary

A questionnaire for determining the present practices of Massachusetts superintendents in regard to measuring teacher efficiency was sent out to 100 localities. The group selected as perhaps representative of the state was chosen by its geographical location and by the 1930 Federal census. The questionnaire included 20 cities, 42 towns with a population of over 5000, and 38 towns under 5000 who maintained high schools. The writer arbitrarily eliminated the three principal cities of Boston, Worcester and Springfield because it was believed that if these cities did have a method, the findings would not be typical of the whole state.

Too much emphasis cannot be placed upon the results of this study for only 76 of the 100 superintendents sent replies and in many cases there was no evidence offered by the superintendent to support his claim. Twenty per cent of the superintendents asserted that they had a systematic and objective method for measuring candidates, and 16 per cent had a similar method for their in-service teachers.

When asked to supply material giving proof of their claim to a method, 16 per cent of them gave some evidence and 26 per cent sent salary schedules ranging in date from 1929 to 1933. Only 9 per cent of the superintendents claimed to have a salary schedule based on merit, and 71 per cent desired a summary of the results of the study.

Forty-eight per cent of the superintendents sent useful information, and the writer's belief that the findings would not be

typical of the whole state if larger cities were included in the study seems justified because 60 per cent of the cities responded with data whereas only 42-43 per cent of the smaller communities furnished any information.

The criteria of school records and training were given by 70 per cent of the superintendents supplying information, for judging prospective candidates, and 64 per cent used references and experience. Only 41 per cent used the interview technique and the writer found three superintendents using the written examination. The implication to be derived from this knowledge and that of the Research Bulletin of the National Education Association in 1932 in which only 3 per cent of the cities in the United States made use of examinations is that educators are satisfied with a candidate's knowledge if he is a graduate of an approved normal school or college.

Butsch's work in gathering together a few studies dealing with the correlation between examination scores and teaching success showed that perhaps tests are invalid as a criteria for judging teacher merit for Knight obtained the highest correlation of +.61 and Tiegs the lowest, +.02, both using the same test. Too much confidence cannot be placed in these findings for the variable of "Judging Teacher Success" was uncontrolled.

Professional study, observation, and rating scales were the chief criteria used by 12 superintendents for judging efficiency of their in-service teachers, and only one mentioned the progress of pupils as the indication of teacher ability. Perhaps those

mentioning observation as a criterion considered pupil progress under that category.

The comments of several superintendents indicate that local political influences govern the selection of teachers in some communities and give evidence that a need exists for a valid and reliable method by which the superintendent can present the qualifications of a teacher without fear of equivocation from his committee.

The salary schedules in many of the smaller communities have become defunct during the economic depression and salaries have been reduced in the larger towns and cities. These reductions have ranged from 5 to 20 per cent. At present the trend is upward for some of the superintendents stated that the salary "cuts" had been restored in part, if not in whole.

The principal criteria for a salary schedule based on teacher merit as given by 7 superintendents was observation and general satisfaction to superintendent and committee, and professional study and successful experience. The meaning to be read into a general statement as "satisfaction to superintendent and committee", and the basis for such a statement being as it is unknown, points out the need for definition of terms and standardization.

The automatic salary increase was mentioned by 33 per cent of the superintendents supplying information and group judgment of teacher ability was mentioned by over 50 per cent of those superintendents who observe their teachers. The comments showed that superintendent, principal, and supervisor acted in this capacity.

The writer gathered 72 different items from the 6 methods presented by Falmouth, Melrose, Longmeadow, North Adams, Andover, and Rockland. These items show only too well how many different conceptions are held by educators as to just what kind or how many qualities should be used as criteria in judging merit. In comparing these 6 methods with the arbitrary criteria listed on page 73 of this study, the writer found that:

1. Personal attributes of teachers were emphasized by 50 per cent of the superintendents.
2. Only one method was a graphic type.
3. The items were not weighted.
4. The five-point evaluation scale was used by 50 per cent of the superintendents.
5. Three of the superintendents used their method as a basis for promotion.
6. The qualities of health, co-operation, and daily preparation had the most frequency of mention.

The writer gathered 72 different items from the 6 methods presented by Falmouth, Melrose, Longmeadow, North Adams, Andover, and Rockland. These items show only too well how many different conceptions are held by educators as to just what kind or how many qualities should be used as criteria in judging merit. In comparing these 6 methods with the arbitrary criteria listed on page 73 of this study, the writer found that:

1. Personal attributes of teachers were emphasized by 50 per cent of the superintendents.
2. Only one method was a graphic type.
3. The items were not weighted.
4. The five-point evaluation scale was used by 50 per cent of the superintendents.
5. Three of the superintendents used their method as a basis for promotion.
6. The qualities of health, co-operation, and daily preparation had the most frequency of mention.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Need for Standardization

Indications of the study.- The findings of this study would indicate that the conditions governing the situation in Massachusetts are typical of the whole teacher measuring movement.

A study of measuring teacher efficiency leaves one with the impression that the number of problems to be encountered and solved would be infinite before a valid and reliable method is produced. No techniques of a statistical nature can possibly obtain the desired results. Clear, critical thought, unbiased and unrelenting in its search for truth, is needed.

The purpose today of measuring a teacher's ability for supervisory functions and self-improvement is the same as when the first method of Meriam's was produced in 1905. Only 10 per cent of those cities having methods for rating teachers use them for other than supervisory practises, and until better methods are devised, salary increments should perhaps be automatic and based upon experience and professional study.

This study found that 68 per cent of the Massachusetts superintendents used the automatic salary schedule with the average yearly increment of 100 dollars. In some of the smaller communities schedules have been abandoned because of financial difficulties due to the economic distress.

Professional study is the principal factor upon which superintendents judge their teachers although not much confidence can be placed in such practice. The courses of study taken by teachers seeking a raise in salary may have no bearing upon the classroom work nor have any value in improving their ability. The only check on such practice is for the superintendent to determine what courses in his opinion would be beneficial to the teachers and advise them along such lines.

The fact that over two billions of dollars was expended in our educational plants in 1930 throughout the United States, and that there is an over-supply of teachers gives indisputable proof of the need for a method which will eliminate those unfit to enter or remain in the profession.

Outstanding Weakness of Methods

The outstanding weakness has not been that the methods which have been presented have had no proven validity and reliability, for such confusion of thought among educators as has been illustrated throughout this study should have warned the reader that validity and reliability are not to be expected.

The methods that have been used have failed principally in their definition and understanding of the terms that should be used for judging teacher efficiency. Educators have agreed that a valid and reliable method is needed but they have allowed confusion of thought to creep into their judgment. Error has been built upon error with the result that the 31 years that have

elapsed since teacher rating came into existence have shown little scientific progress.

Ambiguity of thought and opinion upon the problem has resulted through the uncontrolled, individual efforts of educators, and the interpretation of facts based upon small samplings with variables uncontrolled or incapable of being measured by any present method.

One educator assigns an arbitrary weight to an item on his score card while another bases the weight for the same item upon the normal distribution curve. Both methods are invalid in regard to results obtained and have no scientific value.

Methods have been constructed with no critical analysis of the criteria used; no definition of the meaning of the terms; or no scientific procedure for giving a value, weight, or score to the quality that is measured. The inability to measure objectively traits of personality; to decide scientifically how many items should be on a scale; to discover how many people should do the judging; or the suggestions that teachers' exercises be used as a criterion for teacher measurement; or that the progress of pupils be used as a judgment of teachers, have brought diversity of thought upon the problem instead of crystallization.

The writer believes that these petty, individual efforts of well-meaning educators should be combined under one central research movement.

The first step seems to be that of standardization through a nation-wide experimentation under controlled conditions with

all variables controlled except the teacher factor. The standards set up should be those of universal norms. The time and money needed for such experimentation on so large a scale as indicated, would necessitate an endowment fund from some individual interested in the problem or perhaps government sponsorship. A large corps of educators trained in educational research and a huge clerical force working under central authority would be required. Some of the problems that need to be solved before the foundation of a valid, reliable, and objective method can be produced are as follows:

1. How many items are necessary for measuring validly and reliably a teacher's efficiency?
2. What should be the characteristics of these items or qualities? How defined?
3. How much weight should an item have?
4. When should tests for the achievement quotient technique be given?
5. How long a time should elapse between tests?
6. How many people should judge the merit of a teacher?
7. How many observations and how frequently should they be made?
8. How much of an influence has any particular trait of a teacher upon the reaction of a pupil?
9. Should the graphic scale be used exclusively?
10. How many points should it consist of?

11. Does alternating the extreme values of the criteria on a rating scale eliminate the halo effect of personal bias?

Until such problems are solved and until facilities for conducting such a study are available, no definite progress can be made towards the construction of a valid, reliable, and objective method for measuring teachers in service. As Somers says:

> "Many comprehensive investigations may be required before the major problems....have been solved; but only through experimentation with life situations, as we find them, and with facts as they are, will truth stand revealed, uncontested and incontestable."1/

Recommendations

The Providence plan.--The writer can see only one path to follow at the present time. This is to create a national interest in the problem and to experiment in a scientific manner with such outstanding plans as that now used in the Providence, Rhode Island, school system.

Since there is no valid nor reliable method at present for measuring a teacher's efficiency in service, at least the use of this plan will enable a school system to obtain the best prospects for the positions and may succeed in keeping out the inefficient teacher.

The plan is as follows:

1/ G. T. Somers, "Pedagogical Prognosis: Predicting the Success of Prospective Teachers," New York, Teachers College, Columbia University, 1923, Contributions to Education, No. 140, p.129.

STATEMENT EXPLAINING THE PROVIDENCE PLAN FOR THE OBJECTIVE MEASUREMENT AND RATING OF CANDIDATES FOR TEACHING POSITIONS

Every candidate should possess certain broad educational qualifications including general culture, scholastic aptitude, and professional preparation. All candidates must take objective examinations in these fields. In addition, each candidate should be a master of at least two subject fields of instruction in the curriculum of the public schools and through objective examinations must demonstrate scholarship in these fields sufficient to justify his appointment to teach them. The result of each of these five examinations is translated into a percentile rank, indicating exactly where the candidate would rank in a group of one hundred teachers. These percentile ranks are added and the total examination marks count a possible 500 points out of a total of 1,000.

Naturally the desirability of any system of examinations depends upon the quality, fairness, reliability, method of correction, and method of administration of the tests. Much of the success of the present program is due to the fact that at the time of its inception a new series of tests was made available through the efforts of the American Council on Education and the generosity of the General Education Board. The examinations are of the new-type, objective form, easy to correct, and comparable from year to year. New forms are guaranteed for ten years. The examinations have been prepared by some of the foremost technical experts in the country and are entirely free from commercial motives, since they are available at cost through the foundation.

Three special advantages of these examinations are worthy of note:

1. The objective nature of the examinations makes it possible for each candidate to have the privilege of examining his corrected test to satisfy himself concerning its proper scoring.

2. No teacher who is a college graduate should object to taking the subject examinations that are required of high school and college students.

3. The impartial nature of the examination and its removal from a local setting is highly desirable. Special arrangements insure that no candidate may secure a copy of any test before the date of the examination.

Of the remaining 500 points, the 200 which evaluate training and experience are also impersonal and objective, since they are determined by a scale of values and are based upon a questionnaire which each candidate fills out by mail without his ever visiting

Thus 700 out of 1,000 points are determined on the basis of examinations and facts, all of which may be reviewed and challenged by the candidates and all without involving the names of the candidates.

The remaining 300 points represent a group rating in teaching personality. The rating is made by a group of principals and assistant and deputy superintendents who determine the relative desirability of each candidate as a member of a school faculty, as a wholesome influence with growing young people, and as a positive asset in the life and work of the school. A relative rating on personality is spread on a 300 point scale and added to the ratings on experience, general qualifications, and special qualifications, to obtain a grand total score. When the candidate has been trained in our schools, the record of training is a very important phase of the rating.

When these steps have been taken, the candidates then are listed in subject fields in the order of their total scores and appointed in order as they appear on the list. During the past three years this procedure has been followed. Preference in appointment is given to local candidates. Every candidate is informed concerning his own rating. All records are open for inspection by any person who has a legitimate interest in them, and all candidates who desire further information have been interviewed by the Director of Research or his assistants.

The present plan prevents any suspicion of discrimination on the basis of religion, nationality, race, or political considerations. It places entrance to the teaching profession in Providence upon a high professional plane which considers only the selection of persons best qualified to serve the needs and interests of children and the community. The record of appointments for the past three years speaks for itself in the growing confidence of the teaching staff and of the public in the method and its administration.

The salary increases for teachers are automatic with no stipulated maximum salary. One's natural span of life actually sets the salary limit. A copy of the senior high school schedule illustrates the principle.

Schedule Year	1	2	3	4	5
Schedule Salary	1400	1500	1600	1700	1800
	6	7	8	9	10
	1900	2000	2100	2200	2300
	11	12	13	14	15
	2400	2500	2600	2700	2800

Schedule Year	16	17	18	19	20
Schedule Salary	2900	3000	3000	3100	3100
	21	22	23	24	25
	3100	3200	3200	3200	3200

and so on indefinitely as indicated, with the addition of one year for each succeeding salary level.

APPENDIX

BIBLIOGRAPHY -

Almy, H. C., "A Teacher-Rating Scale of Determined Reliability," Educational Administration and Supervision , 10:182,March 1930.

Armentrout, W. D., "The Rating of Teachers by Training Teachers and Superintendents," Elementary School Journal ,38:511-16, March 1928.

Baird, James and Bates, Guy, "The Basis of Teacher Rating," Educational Administration and Supervision ,15:175-83, March 1929.

Ballou, F.W. "Determining Who Are Superior Teachers," Elementary School Journal , 28:256-62, Dec. 1927.

Barr A. S., An Introduction to the Scientific Study of Classroom Supervision, New York, D.Appleton and Company, 1931, pp.xxv & 399.

_____ Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies ,Bloomington, Illinois, Public School Publishing Company, 1929, pp.viii & 127.

Barr, A. S. and Emans, Lester, "What Qualities Are Prerequisite to Success in Teaching ?", Nations Schools , 6:60-64, Sept. 1930.

Betts, G. L., "Philosophy in the Measurement of Teaching Ability," School Life , 17:168-9, May 1932.

Bonser, F. G., "Qualities Desired in Teachers by Administrators," School and Society , 31:260-62, Feb. 22,1930.

Bossing, Nelson, L., "Teacher-Aptitude Tests and Teacher Selection," Research in Higher Education , Bulletin No. 12, 1931, p.121.

Boyce, A. C., "Methods of Measuring Teachers' Efficiency," Fourteenth Yearbook of the National Society for the Study of Education ,Chicago, University of Chicago Press, 1915, Part II.

Brandenburg, G. C., and Remmers, H. H., "Rating Scales for Instructors," Educational Administration and Supervision , 13:399-406, Sept. 1927.

_____ "Experimental Data on the Purdue Rating Scale for Instructors," Educational Administration and Supervision , 13:519-27, Nov. 1927.

Brown, E. J., "A Self-Rating Scale for Supervisors," American School Board Journal , 77:36-7, August 1928.

Butsch, R. L. G., "Teacher Rating," Review of Educational Research , 1:99-107, 1931.

Bulletin, "A Method for Studying the Unmeasured Traits of an Individual," School of Education, Indiana University, 4:60-73, July 1928.

Chittam, H. D., "The Selection and Appointment of Teachers", Master Thesis, Boston, School of Education, Boston University, 1935.

Collings, Ellsworth, "A Conduct Scale for the Measurement of Teaching," Journal of Educational Method , 6:97-103, Nov. 1926.

Connor, W. L., "A New Method of Rating Teachers," Journal of Educational Research , 1:338-58, May 1920.

Corson, David, B., "The Relation of Scholarship and Growth in Scholarship to Merit and Pay," Addresses and Proceedings of the National Education Association, 59:838, 1921.

Crabbs, Lelah, M., "Measuring Efficiency in Supervision and Teaching," New York, Teachers College, Columbia University, Contribution to Education, No. 175 , 1925.

Day, L. C., "The Teaching Quotient," Elementary School Journal , 33:604-07, April 1933.

Dragoo, A. W., "A Rating Scale for Shop Teachers," Industrial Arts and Vocational Education ,21:8, Jan. 1936.

Eberhart, R., "Evaluating Your Own Teaching," Journal of the National Education Association , 20:304, Nov.1931.

Editorial, "Shall the Teacher-Tenure Laws Be Amended ?", American School Board Journal , 90:43, May 1935.

_____, "Against Stupidity," Boston Sunday Globe , p.4, March 8, 1936.

_____, Boston Globe ,Saturday, p.13, March 7, 1936.

Engelhardt, Max, D., and Tucker, Ledyard, R., "Traits Related to Good and Poor Teaching," School Review , 44:28-33, Jan. 1936.

Engleman, J. O., "The Place of Objective Standards in Judging Student Teaching," Educational Administration and Supervision, 17:321-29, May 1931.

Flam, August, "Rating Industrial-Arts Teachers," Industrial Arts and Vocational Education , 24:63, Feb. 1935.

Flory, Charles, D., "Personality Ratings of Prospective Teachers," Educational Administration and Supervision , 16:135-43, Feb.1930.

Freyd, Max, E., "A Graphic Rating Scale for Teachers," Journal of Educational Research , 8:433-39, Dec. 1933.

Furfey, P. H., "An Improved Rating Scale Technique," Journal of Educational Psychology , 17:45-48, Jan.1926.

Gerling, Henry, J., "Educational Problems," School and Society , 42:101-04, July 20, 1935.

Gilmore, M. E., "Judging and Rating the Teacher," Educational Review , 74:269-73, Dec. 1927.

Goslin, H., "A Salary Schedule for Teachers in a City of 40,000," American School Board Journal ,69:37, Oct. 1924.

Green, Harry, A., and Jorgenson, Albert, N., The Use and Interpretation of High School Tests, New York, Longmans, Green, and Company, 1936, pp. xxvi + 614.

Hart, F. W., Teachers and Teaching, New York, The MacMillan Company, 1934, pp. 285.

Hughes, W., Harding, "General Principles and Results of Rating Trait Characteristics," Journal of Educational Method , 4:421-31, June 1925.

Jacobs, Charles, L., The Relation of the Teacher's Education to Her Effectiveness, New York, Teachers College, Columbia University, Contribution to Education, No. 277 , 1928.

Judd, C. H., Psychology of Secondary Education, Boston, Ginn and Company,1927, pp. xiv + 545.

Knight, F. B., "Qualities Related to Success in Teaching, New York, Teachers College , Columbia University , Contribution to Education No. 120 , 1922.

Knudsen, Charles, W., Evaluation and Improvement of Teaching , New York, Doubleday, Doran and Company, 1932, pp.205-248.

Lancelot, W. H., et al, The Measurement of Teaching Efficiency, New York, The MacMillan Company, 1935, pp. xv + 237.

Lewis, E. E., "Desirable Principles of Teachers' Salary Schedules," American School Board Journal , 80:45, May 1930.

_____, Personnel Problems of the Teaching Staff, New York, The Century Company, 1925, pp. xvii + 460.

Lindsay, E. E., Problems in School Administration , New York, The MacMillan Company, 1928, pp.187-251.

Macdonald, Marion, E., "Students's Opinions as Regards Desirable and Undesirable Qualifications and Practises of Their Teachers in Teacher-Training Institutions," Educational Administration and Supervision , 17:139-46, Feb. 1931.

McCall, William, M., How To Measure in Education, New York,The MacMillan Company, 1927, pp.149-68.

Mead, Arthur, R., Supervised Student-Teaching , New York, Johnson Publishing Company, 1930, pp.467-512.

Morris, E. H., Personal Traits and Success in Teaching , New York, Teachers College, Columbia University, Contribution to Education No. 342, 1929.

Morrison, Henry, C., The Practise of Teaching in the Secondary School, Chicago, University of Chicago Press, 1931, pp.121-40.

Mort, Paul, R., "State Support for Public Education," National Survey of United States School Finance, Biennial Survey,United States Office of Education,1933, p.309.

Nanninga, S. P., "Estimates of Teachers in Service," School Review, 36:622-26, Oct. 1928.

National Education Association Bulletin, Research Bulletin,"The Status of Teacher Demand and Supply in the Elementary and Secondary Schools," 9:338-49, Nov. 1931.

_____, Research Bulletin, "Administrative Practises Affecting Classroom Teachers," 10:115, March 1932.

_____, First Yearbook,Department of Classroom Teachers, pp.45-50, July 1926.

National Society for the Study of Education, Twelfth Yearbook, "The Supervision of City Schools," Bloomington,Illinois, Public School Publishing Company, 1923, pp.7-96.

National Survey of the Education of Teachers, "Special Survey Studies" Vol.1,No. 10, Office of Education, United States Department of the Interior,1933,pp.117-153.

Nichols, A. M., "Value and Procedure in Using Schemes for Evaluating Student Teachers," Master Thesis ,Boston, School of Education, Boston University, 1932.

Newlon, Jesse, H., "Salary Differences, If Any, Based on the Rank or Class of the Teacher," Addresses and Proceedings of the National Education Association, 59:841, 1921.

Newmark, David, "Students' Opinions of Their Best and Poorest Teachers" Elementary School Journal, 29:576-85, April 1929.

Nutt, H. W., The Supervision of Instruction , Boston, Houghton Mifflin Company, 1920, pp.214-29.

Ruediger, W. C., "Rating Teachers," School and Society, 20:266, August 30, 1924.

Rugg, H. O., "Self-Improvement of Teachers Through Self-Rating," Elementary School Journal, 19:670-84, May 1920.

Somers, G. T., "Pedagogical Prognosis: Predicting the Success of Prospective Teachers, New York, Teachers College, Columbia University, Contribution to Education, No. 140, 1923.

Sorenson, Herbert, "Why Teaching Success Does not Correlate Highly with Measured Intelligence," Educational Administration and Supervision, 15:602-06, Nov. 1929.

Starrak, J. A., "The Construction of a Scale for the Measurement of College Teaching, and the Determination of its Reliability and Validity," Doctor Dissertation, Boston, School of Education, Boston University, 1932.

Strebel, R. F., "Purposes of Evaluation of Student Teaching," Educational Administration and Supervision, 17:336-40, May 1931.

Thurstone, L. L., "Attitudes Can Be Measured," American Journal of Sociology, 33:529-54, Jan. 1928.

Towner, Earl, M., "The National Elementary Principal," Bulletin of the Department of Elementary School Principals of the National Education Association, 15:14-16, Feb. 1936.

Weber, Oscar, F., "Problems in Public School Administration," New York, The Century Company, 1930, pp.xxxviii & 726.

Wetzel, W. A., "Plan for Rating Teachers Based Upon Pupil Accomplishment," School Life, 13:26-27, Oct. 1927.

Williams, Eula, S., "A Personality Rating Form for Elementary School Pupils," Elementary School Journal, 34:16-29, Sept. 1933.

BOSTON UNIVERSITY
1 1719 02548 6301